

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 9, 2005, 02:21:19 ; Search time 3730 Seconds
(without alignments)
9735.473 Million cell updates/sec

Title: US-09-087-136-2

Perfect score: 954
Sequence: 1 atgtcagaatagatccacac.....ttcaatcacatcttgac 954

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_est3: *
4: gb_est4: *
5: gb_est5: *
6: gb_est6: *
7: gb_est7: *
8: gb_est8: *
9: gb_est9: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	727.4	76.2	752	1	AU218734 AU218734
2	567.2	59.5	666	6	CB102172 MMV_SQ005
3	559	58.6	559	6	CB386085 OSTF038D2
4	532	55.8	614	6	CB102170 MMV_SQ005
5	506	53.0	605	6	CB102173 MMV_SQ005
6	497	52.1	584	6	CB102177 MMV_SQ005
7	489	51.3	586	6	CB102175 MMV_SQ005
8	483	50.6	562	6	CB102169 MMV_SQ005
9	411	43.1	495	6	CB102055 MMV_SQ001
10	397.2	41.6	485	6	CB102059 MMV_SQ001
11	365	38.3	377	1	AV202470 AV202470
12	354	37.1	360	7	D74918 D74918
13	354	37.1	360	7	D75085 D75085
14	354	37.1	360	7	CB102066 MMV_SQ001
15	353	37.0	433	6	CB102067 MMV_SQ001
16	347	36.4	425	6	CB102064 MMV_SQ001
17	346	36.3	425	6	CB102065 MMV_SQ001
18	334.4	35.1	360	7	D69562 D69562
19	330.8	34.7	360	7	AV203242 AV203242
20	314.4	33.0	339	7	D65944 D65944
21	300	31.4	300	6	C55654 C55654
22	297.4	31.2	300	6	CB102062 MMV_SQ001
23	297.4	30.6	300	1	AV199733 AV199733
C 24	291.8	30.6	300	1	AV200339 AV200339

C 25	173.8	18.2	180	7	D72252 D72252
26	65.2	6.8	653	6	CA958252 CA958252
27	48.8	5.1	1101	9	CNS00395 CNS00395
C 28	47.2	4.9	405	1	AI447427 AI447427
29	45.6	4.8	392	2	BF661811 BF661811
C 30	45.6	4.8	427	1	AI843070 AI843070
31	45.6	4.8	438	1	AA163826 AA163826
C 32	45.6	4.8	461	5	BX519649 BX519649
C 33	45.6	4.8	734	5	BO177189 BO177189
34	45.4	4.7	948	8	BH148509 BH148509
35	45	4.7	435	1	AI592053 AI592053
C 36	44.8	4.7	641	1	AV735039 AV735039
C 37	44.2	4.6	891	9	CNS01011 CNS01011
C 38	44	4.6	296	2	BB247250 BB247250
C 39	43.2	4.5	935	8	AZ672288 AZ672288
40	43.2	4.5	1101	9	CNS00380 CNS00380
C 41	43	4.5	448	5	BQ857965 BQ857965
C 42	43	4.5	1000	9	CNS00000 CNS00000
C 43	43	4.5	1101	9	CNS00033 CNS00033
C 44	43	4.5	1101	9	CNS01611 CNS01611
C 45	42.8	4.5	982	8	CC107246 CC107246

ALIGNMENTS

RESULT 1	AU218734	752 bp	mRNA	linear	EST 17-JUL-2001
LOCUS	AU218734	unpublished oligo-capped cDNA library, stage 11			
DEFINITION	AU218734	unpublished oligo-capped cDNA library, stage 11			
ACCESSION	AU218734	unpublished oligo-capped cDNA clone yk686e01 3', mRNA sequence.			
VERSION	AU218734				
KEYWORDS	AU218734.1	GI:14856891			
SOURCE	EST.				
ORGANISM	Caenorhabditis elegans				
REFERENCE	Caenorhabditis elegans				
AUTHORS	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.				
TITLE	1 (bases 1 to 752)				
JOURNAL	Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y. and Sugano, S.				
COMMENT	A complementary view of the C.elegans genome				
UNPUBLISHED (2001)	Contact: Yui Kohara				
Genome Biology Lab.					
National Institute of Genetics					
Yata 1111, Mishima, Shizuoka 411, Japan					
Tel: 81-559-81-6854					
Fax: 81-559-81-6855					
Email: ykohara@lab.nig.ac.jp.					
Location/Qualifiers					
1. 752					
/organism="Caenorhabditis elegans"					
/mol_type="mRNA"					
/db_xref="taxon:6239"					
/clone="YK686e01"					
/sex="Hermaphrodite"					
/tissue_type="whole animal"					
/dev_stage="L1"					
/clone_11b="unpublished oligo-capped cDNA library, stage 11"					
ORIGIN					
Query Match	76.2%; Score 727.4; DB 1; Length 752;				
Best Local Similarity	99.6%; Pred. No. 7.2e-181;				
Matches	750; Conservative 0; Mismatches 1; Indels 2; Gaps 2;				
CY	195 AATCTGCTGACAA-CATTAACATCACCACATCGGAAATTAATTTGGAGTACTCCAGAG 253				
DB	752 AATCTGCTGACAA-CATTAACATCACCACATCGGAAATTAATTTGGAGTACTCCAGAG 693				
CY	254 AATTAATCTGCTGATGAAGAGCCGGGCGCCAGGACGTCCAGCGTCCGCTCTGTA 313				

Db 692 ATTATACGTCGTCGATGAGAGCCCGGCGCAAGAGCTCCACCAAGTGGCCCTCGTA 633
Qy 314 AGATCCCGCGTCACGAATCTTCACTTATGGAATCACCAGCAAGACTATGATC 373
Db 632 AGATCCCGCGTCACGAATCTTCACTTATGGAATCACCAGCAAGACTATGATC 573
Qy 374 GTGATTTCAATATATGTTGAAATGCGTGAAAAACATTCGAATGATGAGTGAAGT 433
Db 572 GTGATTTCAATATATGTTGAAATGCGTGAAAAACATTCGAATGATGAGTGAAGT 513
Qy 434 TTGAAGAAATATATGTTGAAATGCGTGAAAAACATTCGAATGATGAGTGAAGT 493
Db 512 TTGAAGAAATATATGTTGAAATGCGTGAAAAACATTCGAATGATGAGTGAAGT 453
Qy 494 AACCAATTAAGAGCTCAAGAGACAGCTATGACCCGAACTGGCTGTGATTTCTGAT 553
Db 452 AACCAATTAAGAGCTCAAGAGACAGCTATGACCCGAACTGGCTGTGATTTCTGAT 393
Qy 554 GTGCGAGATTCATCGAATGCCAGCTCCAGATTAATTCCTGAGCTGCCAATGTTTC 613
Db 392 GTGCGAGATTCATCGAATGCCAGCTCCAGATTAATTCCTGAGCTGCCAATGTTTC 333
Qy 614 CATCTAGAATCGATGAATTCGACGCTACAGTCGATCCAAATATGAAAAGATTTGAAA 673
Db 332 CATCTAGAATCGATGAATTCGACGCTACAGTCGATCCAAATATGAAAAGATTTGAAA 273
Qy 674 ATGAATATATCTGCTATGGAACAAAGTCAAAAAAGTTGGTGCTCTATCAAGCTGTC 733
Db 272 ATGAATATATCTGCTATGGAACAAAGTCAAAAAAGTTGGTGCTCTATCAAGCTGTC 213
Qy 734 GAGCTGCTCCCATGCAAGAGCATGAGTATTAATCAAAATCTACAGCTGGAGAGT 793
Db 212 GAGCTGCTCCCATGCAAGAGCATGAGTATTAATCAAAATCTACAGCTGGAGAGT 153
Qy 794 CGAAAAGCTGCGAGACCACTGGCTTATTTAAATATCTTAACATGTAATTTCAATTT 853
Db 152 CGAAAAGCTGCGAGACCACTGGCTTATTTAAATATCTTAACATGTAATTTCAATTT 93
Qy 854 ATCTCTTATCTTCTGATCTTGTATCAATGCTCTTATTTCAAAAACTCACTTTAAA 913
Db 92 ATCTCTTATCTTCTGATCTTGTATCAATGCTCTTATTTCAAAAACTCACTTTAAA 33
Qy 914 TTCATATTAATATGAGTTTATTTCAATATCATC 946
Db 32 TTCATATTAATATGAGTTTATTTCAATATCATC 1

RESULT 2
CB102172 666 bp mRNA linear EST 28-JAN-2003
LOCUS Mv1500586 AD-wrmCDNA Caenorhabditis elegans cDNA, mRNA sequence.
DEFINITION CB102172.1 GI:27927979
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Caenorhabditis elegans
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.
REFERENCE
AUTHORS
Walhout, A.J., Sordeja, R., Ju, X., Hartley, J.L., Temple, G.F.,
Bresch, M.A., Thierry-Mieg, N., and Vidal, M.
TITLE
Protein interaction mapping in C. elegans using proteins involved
in vulval development
JOURNAL
Science 287 (5450), 116-122 (2000)
MEDLINE
PUBMED
COMMENT
Contact: Vidal, M.
Marc Vidal Laboratory
Dana Farber Cancer Institute
1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
Tel: 617 632 5180
Fax: 617 632 5739

FEATURES
source
Location/Qualifiers
1..666
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/sex="hermaphrodite and male"
/tissue_type="whole animal"
/dev_stage="mixed stage"
/clone_id="AD-wrmCDNA"
/note="The AD-wrmCDNA library was generated with poly(A)+
RNA isolated from both hermaphrodite and male N2 worms of
all larval stages, embryos, adults and dauers and the
subsequent generation of cDNAs by poly(A) priming. The
cDNAs were cloned into pC86"

ORIGIN
Query: Match 59.5%; Score 567.2; DB 6; Length 666;
Best Local Similarity 99.5%; Pred. No. 1.4e-138;
Matches 569; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 ATGTCAAGAAATAGATCCACTTGGCCGAGTCTTCTCCAGAAAGCGAGATCGAAATGCT 60
Db 95 ATGTCAAGAAATAGATCCACTTGGCCGAGTCTTCTCCAGAAAGCGAGATCGAAATGCT 154
Qy 61 CGTCAAAATGATCCATGATTAAGCGAGGTCCTTCCATTGGAATGCCAAGCAAAA 120
Db 155 CGTCAAAATGATCCATGATTAAGCGAGGTCCTTCCATTGGAATGCCAAGCAAAA 214
Qy 121 CTGACATCCCTGTTATCTATGATCCGACAGTTCCCGAGTCCCGATATGAATTTGCC 180
Db 215 CTGACATCCCTGTTATCTATGATCCGACAGTTCCCGAGTCCCGATATGAATTTGCC 274
Qy 181 AGAAAAGCTGGAATCTGCTGACCAACCATTAACATCAACCATCGAAATTAATGGA 240
Db 275 AGAAAAGCTGGAATCTGCTGACCAACCATTAACATCAACCATCGAAATTAATGGA 334
Qy 241 GATCTCCGAAGATTAATCTGCTGATGAAGAGCCGCGGCGCAAGAGCTGCCACA 300
Db 335 GATCTCCGAAGATTAATCTGCTGATGAAGAGCCGCGGCGCAAGAGCTGCCACA 394
Qy 301 GGTGCGCTGTAAGATGCGCGGTCAGCAATCTTCAATCTCACTTATGGAATGACACGC 360
Db 395 GGTGCGCTGTAAGATGCGCGGTCAGCAATCTTCAATCTCACTTATGGAATGACACGC 454
Qy 361 AAGACTATGACTGCTGATTTCAAAATTAATGTTGAATTTGCGTGAAGAAACCATTCGAATG 420
Db 455 AAGACTATGACTGCTGATTTCAAAATTAATGTTGAATTTGCGTGAAGAAACCATTCGAATG 514
Qy 421 ATAGCTGACGTTTGAAGAAATATCTTCTGTTAGACATGCGTTAAAGACACATG 480
Db 515 ATAGCTGACGTTTGAAGAAATATCTTCTGTTAGACATGCGTTAAAGACACATG 574
Qy 481 AATATGAATATGACCAATTAAGAGCTCAAGAGACAGATGACCGAATCTGCGTGT 540
Db 575 AATATGAATATGACCAATTAAGAGCTCAAGAGACAGATGACCGAATCTGCGTGT 634
Qy 541 GATTATCTTGCATGTCGCGAGATTCATCGAAT 572
Db 635 GATTATCTTGCATGTCGCGAGATTCATCGAAT 666

RESULT 3
CB386085 559 bp mRNA linear EST 15-MAY-2003
LOCUS OSTF038D2_1 AD-wrmCDNA Caenorhabditis elegans cDNA, mRNA sequence.
DEFINITION CB386085
ACCESSION
VERSION
KEYWORDS
SOURCE
Caenorhabditis elegans

ORGANISM *Caenorhabditis elegans*
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; *Caenorhabditis*.
 1 (bases 1 to 559)
 AUTHORS Rebou, J., Vaglio, P., Rual, J.F., Lamesch, P., Martinez, M.,
 Armstrong, C.M., Li, S., Jacotot, L., Bertin, N., Janky, R., Moore, T.,
 Hudson, J.R., Hartley, J.L., Brasch, M.A., Vandenhaute, J., Boulton, S.,
 Endress, G.A., Jenna, S., Chevet, E., Papadimitropoulos, V.,
 Tojias, P.P., Pracek, J., Snyder, M., Huang, R., Chance, M.R., Lee, H.,
 Doucette-Stamm, L., Hill, D.E. and Vidal, M.
 C. elegans ORFome version 1.1: experimental verification of the
 genome annotation and resource for proteome-scale protein
 expression
 Nat. Genet. (2003) In press
 Contact: Vidal, M.
 JOURNAL Marc Vidal Laboratory
 COMMENT Dana Farber Cancer Institute
 1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
 Tel: 617 632 5180
 Fax: 617 632 5739
 Email: Marc.Vidal@dfci.harvard.edu
 Sequence tag of Gateway entry clones. The primers used were
 designed on the predicted protein encoding ORF. C. elegans ORFome
 cloning project : Contact david_hill@dfci.harvard.edu or
 marc.vidal@dfci.harvard.edu
 POLYA=No.

FEATURES
 source
 1..559
 /location/Qualifiers
 /organism="Caenorhabditis elegans"
 /mol_type="mRNA"
 /strain="N2"
 /db_xref="taxon:6239"
 /sex="Hermaphrodite and male"
 /tissue_type="whole animal"
 /dev_stage="mixed stage"
 /clone_1lb="AD-wrmCDNA"
 /note="The AD-wrmCDNA library was generated with poly(A)+
 RNA isolated from both hermaphrodite and male N2 worms of
 all larval stages, embryos, adults and dauers and the
 subsequent generation of cDNAs by poly(A) priming. The
 cDNAs were cloned into pPC86"

ORIGIN
 Query Match 58.6%; Score 559; DB 6; Length 559;
 Best Local Similarity 100.0%; Pred. No. 2e-136;
 Matches 559; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

G
 25 GAGTTCTTGCTTCAGAAAGACGAGATGCAATGCTGCTCAAAATGATCATTGATAAGC 84
 |||||
 D 1 GAGTTCTTGCTTCAGAAAGACGAGATGCAATGCTGCTCAAAATGATCATTGATAAGC 60
 |||||

G
 85 GAGAGTCCACTTCATTGGAATGCGCAAGCAAGAAATCAGATCCCTGTTATCTTATGAT 144
 |||||
 D 61 GAGAGTCCACTTCATTGGAATGCGCAAGCAAGAAATCAGATCCCTGTTATCTTATGAT 120
 |||||

G
 145 CCGACAGTCCGGAGTCAACCGGATATGAATTCGCCAGAAAGCGTCTGGGAAATCTGTG 204
 |||||
 D 121 CCGACAGTCCGGAGTCAACCGGATATGAATTCGCCAGAAAGCGTCTGGGAAATCTGTG 180
 |||||

G
 205 ACAACATTAACATCAACCCATCGAAATTAATGAGTACCTCCAGAAATTTATCTGCT 264
 |||||
 D 181 ACAACATTAACATCAACCCATCGAAATTAATGAGTACCTCCAGAAATTTATCTGCT 240
 |||||

G
 265 GCTGATGAAGAGCCCGGGCGCCAGAGAGTCCACAGGTGCGCTCGTAAGATCCCGGT 324
 |||||
 D 241 GCTGATGAAGAGCCCGGGCGCCAGAGAGTCCACAGGTGCGCTCGTAAGATCCCGGT 300
 |||||

G
 325 CAGCAATTTCAACTTCACTTATGGAATCAGCAAGCAAGATATGATCGCGATTTCTTAA 384
 |||||
 D 301 CAGCAATTTCAACTTCACTTATGGAATCAGCAAGCAAGATATGATCGCGATTTCTTAA 360
 |||||

G
 385 ATTATGTTGAATTCGTGAGAAACATTCGAAATGATGATGCGTTTGAAGAGAA 444
 |||||

Db 361 ATTATGTTGAATTCGTGAGAAACATTCGAAATGATGATGCGTTTGAAGAGAA 420
 |||||
 G 445 TATTCACCTTGTTAGAGCATGGGTTTAAAGACATGATTAATGATTAACCAATATAA 504
 |||||
 D 421 TATTCACCTTGTTAGAGCATGGGTTTAAAGACATGATTAATGATTAACCAATATAA 480
 |||||

G
 505 GCTCAAAAGACAGACTATGACCGCAATCTGCGTGTATTTCTTTCATGTCGAGATT 564
 |||||
 D 481 GCTCAAAAGACAGACTATGACCGCAATCTGCGTGTATTTCTTTCATGTCGAGATT 540
 |||||

G
 565 CATGAATGCCACGCTCCAG 583
 |||||
 D 541 CATGAATGCCACGCTCCAG 559
 |||||

RESULT 4
 CBI02170 614 bp mRNA linear EST 28-JAN-2003
 LOCUS MWV S00050B4 AD-wrmCDNA *Caenorhabditis elegans* cDNA, mRNA sequence.
 DEFINITION CBI02170
 ACCESSION CBI02170.1 GI:27927977
 VERSION
 KEYWORDS EST.
 SOURCE
 ORGANISM *Caenorhabditis elegans*
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; *Caenorhabditis*.
 1 (bases 1 to 614)
 Walhout, A.J., Sordella, R., Lu, X., Hartley, J.L., Temple, G.F.,
 Brasch, M.A., Thierry-Mieg, N. and Vidal, M.
 Protein interaction mapping in C. elegans using proteins involved
 in vulval development
 Science 287 (5450), 116-122 (2000)
 JOURNAL MEDLINE
 20082953
 PUBMED 10615043
 COMMENT Contact: Vidal M
 Marc Vidal Laboratory
 Dana Farber Cancer Institute
 1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
 Tel: 617 632 5180
 Fax: 617 632 5739
 Email: Marc.Vidal@dfci.harvard.edu
 Trace dvp10316.x with Bait 1in-53
 POLYA=No.

FEATURES
 source
 1..614
 /location/Qualifiers
 /organism="Caenorhabditis elegans"
 /mol_type="mRNA"
 /strain="N2"
 /db_xref="taxon:6239"
 /sex="Hermaphrodite and male"
 /tissue_type="whole animal"
 /dev_stage="mixed stage"
 /clone_1lb="AD-wrmCDNA"
 /note="The AD-wrmCDNA library was generated with poly(A)+
 RNA isolated from both hermaphrodite and male N2 worms of
 all larval stages, embryos, adults and dauers and the
 subsequent generation of cDNAs by poly(A) priming. The
 cDNAs were cloned into pPC86"

ORIGIN
 Query Match 55.8%; Score 532; DB 6; Length 614;
 Best Local Similarity 100.0%; Pred. No. 2.7e-129;
 Matches 532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

G
 5 CAGAAATAGATTCACCTTGCGAGTTCTTGCTTCAGAAAGACGAGATGCAATGCTGCTC 64
 |||||
 D 83 CAGAAATAGATTCACCTTGCGAGTTCTTGCTTCAGAAAGACGAGATGCAATGCTGCTC 142
 |||||

G
 65 AAAATGATTCATGATAGAGCGAGAGTCCACTTCATTGGAATGCCCAAGCAAGAAATCTCA 124
 |||||
 D 143 AAAATGATTCATGATAGAGCGAGAGTCCACTTCATTGGAATGCCCAAGCAAGAAATCTCA 202
 |||||

G
 125 CATCCCTTATCTTATGATCCGACAGTCCGAGATCAACCGGATATGAATTCGCCAGAA 184
 |||||

Db 203 CATCCCTGTTATCTTATGATCCGACAGTCCGAGATGACCGGATATGAAATTCGCCAGAA 262
 Qy 185 AACGCTGGGAATCTGCTGACCAACATAAAACATCACCCATCGGAATTAATGGAGTAC 244
 Db 263 AACGCTGGGAATCTGCTGACCAACATAAAACATCACCCATCGGAATTAATGGAGTAC 322
 Qy 245 TCCCAAGATTAATCTGCTGATGAGAGCCCGGCGCCAGACGTCACAGTGC 304
 Db 323 TCCCAAGATTAATCTGCTGATGAGAGCCCGGCGCCAGACGTCACAGTGC 382
 Qy 305 GCCCTCGTAAGATGCGCGCTCAAGATCTTCACTTCACTTATGGAATCACACGCAAGA 364
 Db 383 GCCCTCGTAAGATGCGCGCTCAAGATCTTCACTTCACTTATGGAATCACACGCAAGA 442
 Qy 365 CTATGACTGCTGATCTTAAATTAATGTTGAATTCGCGGAAACCATTCGAATGATAG 424
 Db 443 CTATGACTGCTGATCTTAAATTAATGTTGAATTCGCGGAAACCATTCGAATGATAG 502
 Qy 425 CTGGAAGTTTGAAGAAATATGACTTGTGAGAGCATGGGTTAAAGGACATGAAATA 484
 Db 503 CTGGAAGTTTGAAGAAATATGACTTGTGAGAGCATGGGTTAAAGGACATGAAATA 562
 Qy 485 ATGAATATGACCAATTAAGCTCAAGACGACTATGACCCGATCTGGC 536
 Db 563 ATGAATATGACCAATTAAGCTCAAGACGACTATGACCCGATCTGGC 614
 RESULT 5
 CBI02173 605 bp mRNA linear EST 28-JAN-2003
 LOCUS MWV SQ005@B8 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA sequence.
 DEFINITION CBI02173
 ACCESSION CBI02173.1 GI:27927980
 VERSION EST.
 KEYWORDS Caenorhabditis elegans
 SOURCE Caenorhabditis elegans
 ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditioidea; Rhabditidae; Pelodertinae; Caenorhabditis.
 1 (bases 1 to 605)
 Walhout,A.J., Sordella,R., Lu,X., Hartley,J.L., Temple,G.F.,
 Brasch,M.A., Thierry-Mieg,N. and Vidal,M.
 Protein interaction mapping in C. elegans using proteins involved
 in vulval development
 Science 287 (5450), 116-122 (2000)
 JOURNAL MEDLINE
 PUBMED 10615043
 COMMENT Contact: Vidal M
 Marc Vidal Laboratory
 Dana Farber Cancer Institute
 1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
 Tel: 617 632 5180
 Fax: 617 632 5739
 Email: Marc.Vidal@dfci.harvard.edu
 Trace dyp10320.x with Bait 1in-53
 POLYA-No.
 FEATURES
 source 1..605
 /organism="Caenorhabditis elegans"
 /mol_type="mRNA"
 /strain="N2"
 /db_xref="taxon:6239"
 /sex="Hermaphrodite and male"
 /tissue_type="whole animal"
 /dev_stage="mixed stage"
 /clone_id="AD-wrmcDNA"
 /note="The AD-wrmcDNA library was generated with poly(A)+
 RNA isolated from both hermaphrodite and male N2 worms of
 all larval stages, embryos, adults and dauers and the
 subsequent generation of cDNAs by poly(A) priming. The
 cDNAs were cloned into pPC66"

Query Match 53.0%; Score 506; DB 6; Length 605;
 Best Local Similarity 100.0%; Pred. No. 2e-122;
 Matches 506; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 ATGTCAGAAATAGATCCATTGCTTCTTCTTCCAGAAAGACGAGATCGAAATCT 60
 Db 100 ATGTCAGAAATAGATCCATTGCTTCTTCTTCCAGAAAGACGAGATCGAAATCT 159
 Qy 61 CGTCAAAATGATTCATTTGATTAAGCGAGGTCACCTTCATTGGAAATCGCAAGAGAAA 120
 Db 160 CGTCAAAATGATTCATTTGATTAAGCGAGGTCACCTTCATTGGAAATCGCAAGAGAAA 219
 Qy 121 CTACATCCCTGTTATCTTATGATCCGACAGTTCGAGATCCGAGATGAAATTCGCC 180
 Db 220 CTACATCCCTGTTATCTTATGATCCGACAGTTCGAGATCCGAGATGAAATTCGCC 279
 Qy 181 AGAAAAGCTGTGGAAATCTGCTGACACCATTAATCAACCCATCGGAATTAATGGA 240
 Db 280 AGAAAAGCTGTGGAAATCTGCTGACACCATTAATCAACCCATCGGAATTAATGGA 339
 Qy 241 GTACTCCCAAGAAATTAATCTGCTGATGAGAGCCCGGCGCCAGACGTCACCA 300
 Db 340 GTACTCCCAAGAAATTAATCTGCTGATGAGAGCCCGGCGCCAGACGTCACCA 399
 Qy 301 GGTGCGCCTCTGTAAGATGCGCGCTCAAGATCTTCACTTCACTTATGGAATCACAGGC 360
 Db 400 GGTGCGCCTCTGTAAGATGCGCGCTCAAGATCTTCACTTCACTTATGGAATCACAGGC 459
 Qy 361 AAGACTTGAATCTGATTTCTAAATTAATGTTGAATTCGCGGAAACCATTCGAATG 420
 Db 460 AAGACTTGAATCTGATTTCTAAATTAATGTTGAATTCGCGGAAACCATTCGAATG 519
 Qy 421 ATAGCTGAGCGTTTGAAGAAATATTCATTGTTGAGAGCATGGGTTAAAGACATG 480
 Db 520 ATAGCTGAGCGTTTGAAGAAATATTCATTGTTGAGAGCATGGGTTAAAGACATG 579
 Qy 481 AATATGAAATATGAAACCAATTAAGC 506
 Db 580 AATATGAAATATGAAACCAATTAAGC 605
 RESULT 6
 CBI02177 584 bp mRNA linear EST 28-JAN-2003
 LOCUS MWV SQ005@B12 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA
 DEFINITION CBI02177
 ACCESSION CBI02177.1 GI:27927984
 VERSION EST.
 KEYWORDS Caenorhabditis elegans
 SOURCE Caenorhabditis elegans
 ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditioidea; Rhabditidae; Pelodertinae; Caenorhabditis.
 1 (bases 1 to 584)
 Walhout,A.J., Sordella,R., Lu,X., Hartley,J.L., Temple,G.F.,
 Brasch,M.A., Thierry-Mieg,N. and Vidal,M.
 Protein interaction mapping in C. elegans using proteins involved
 in vulval development
 Science 287 (5450), 116-122 (2000)
 JOURNAL MEDLINE
 PUBMED 10615043
 COMMENT Contact: Vidal M
 Marc Vidal Laboratory
 Dana Farber Cancer Institute
 1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
 Tel: 617 632 5180
 Fax: 617 632 5739
 Email: Marc.Vidal@dfci.harvard.edu
 Trace dyp10324.x with Bait 1in-53
 POLYA-No.
 FEATURES
 source 1..584
 /organism="Caenorhabditis elegans"

/mol_type="mRNA"
 /strain="N2"
 /db_xref="taxon:6239"
 /sex="Hermaphrodite and male"
 /tissue_type="whole animal"
 /dev_stage="mixed stage"
 /clone_id="AD-wrmcDNA"
 /note="The AD-wrmcDNA library was generated with poly(A)+ RNA isolated from both hermaphrodite and male N2 worms of all larval stages, embryos, adults and dauers and the subsequent generation of cDNAs by poly(A) priming. The cDNAs were cloned into pPC86"

ORIGIN

Query Match 52.1%; Score 497; DB 6; Length 584;
 Best Local Similarity 100.0%; Pred. No. 4.6e-120;
 Matches 497; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTCAGAAATAGATCCACTTGGCGAGTCTTGTCCAGAAAGCGGAGATCGAAATGCT 60
 DB 8 ATGTCAGAAATAGATCCACTTGGCGAGTCTTGTCCAGAAAGCGGAGATCGAAATGCT 147
 QY 61 CGTCMAAATGATCCATTGATTAAGCGAGTCCACTTCCATTGGAATCGCCAAAGCAGAAA 120
 DB 148 CGTCMAAATGATCCATTGATTAAGCGAGTCCACTTCCATTGGAATCGCCAAAGCAGAAA 207
 QY 121 CTCACATCCCTGTTATCTTATGATCCGACAGTTCGCGAGTCCAGGATATGAATTTGCC 180
 DB 208 CTCACATCCCTGTTATCTTATGATCCGACAGTTCGCGAGTCCAGGATATGAATTTGCC 267
 QY 181 AGAAAGCGTCGGGAAATCTGCTGACAAACATTAACATCCAGTCCGAAATTAATGGA 240
 DB 268 AGAAAGCGTCGGGAAATCTGCTGACAAACATTAACATCCAGTCCGAAATTAATGGA 327
 QY 241 GTACTCCAGAAAGTATTAATCTGCTGATGAGAGCCGCGGCCCAAGAGCGTCCACA 300
 DB 328 GTACTCCAGAAAGTATTAATCTGCTGATGAGAGCCGCGGCCCAAGAGCGTCCACA 387
 QY 301 GGTGCGCCCTGTAAGATGCGCGCTGACGAACTTCACTTATGGAATCAACACGC 360
 DB 388 GGTGCGCCCTGTAAGATGCGCGCTGACGAACTTCACTTATGGAATCAACACGC 447
 QY 361 AAGACTATGACTCGTGAATCTCAAAATTAATGTTGAATGCGGAGAAACCATTTGAATG 420
 DB 448 AAGACTATGACTCGTGAATCTCAAAATTAATGTTGAATGCGGAGAAACCATTTGAATG 507
 QY 421 ATAGCTGACGCTTTTGAAGAAGATTAATCACTTGGTAGAGCATGGTTAAAGACACATG 480
 DB 508 ATAGCTGACGCTTTTGAAGAAGATTAATCACTTGGTAGAGCATGGTTAAAGACACATG 567
 QY 481 AATAATGAAATGTAAC 497
 DB 568 AATAATGAAATGTAAC 584

RESULT 7
 CBI02175 586 bp mRNA linear EST 28-JAN-2003
 LOCUS MW_50005@B10 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA
 DEFINITION sequence.
 ACCESSION CBI02175
 VERSION CBI02175.1 GI:27927982
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans
 ORGANISM Caenorhabditis elegans
 BUKARYOTA; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 586) Sordella, R., Lu, X., Hartley, J. L., Temple, G. F.,
 Walhout, A. J., Thierry-Mieg, N. and Vidal, M.
 Protein interaction mapping in C. elegans using proteins involved
 in vulval development
 JOURNAL Science 287 (5450), 116-122 (2000)

MEDLINE 20082953
 PUBMED 10615043
 COMMENT Contact: Vidal M
 Marc Vidal Laboratory
 Dana Farber Cancer Institute
 1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
 Tel: 617 632 5180
 Fax: 617 632 5729
 Email: Marc.Vidal@dfci.harvard.edu
 Trace dvp10322.x with Bait 11n-53
 POLYA=No.

FEATURES

source Location/Qualifiers
 1..586
 /organism="Caenorhabditis elegans"
 /mol_type="mRNA"
 /strain="N2"
 /db_xref="taxon:6239"
 /sex="Hermaphrodite and male"
 /tissue_type="whole animal"
 /dev_stage="mixed stage"
 /clone_id="AD-wrmcDNA"
 /note="The AD-wrmcDNA library was generated with poly(A)+ RNA isolated from both hermaphrodite and male N2 worms of all larval stages, embryos, adults and dauers and the subsequent generation of cDNAs by poly(A) priming. The cDNAs were cloned into pPC86"

ORIGIN

Query Match 51.3%; Score 489; DB 6; Length 586;
 Best Local Similarity 100.0%; Pred. No. 6e-118;
 Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTCAGAAATAGATCCACTTGGCGAGTCTTGTCCAGAAAGCGGAGATCGAAATGCT 60
 DB 98 ATGTCAGAAATAGATCCACTTGGCGAGTCTTGTCCAGAAAGCGGAGATCGAAATGCT 157
 QY 61 CGTCMAAATGATCCATTGATTAAGCGAGTCCACTTCCATTGGAATCGCCAAAGCAGAAA 120
 DB 158 CGTCMAAATGATCCATTGATTAAGCGAGTCCACTTCCATTGGAATCGCCAAAGCAGAAA 217
 QY 121 CTCACATCCCTGTTATCTTATGATCCGACAGTTCGCGAGTCCAGGATATGAATTTGCC 180
 DB 218 CTCACATCCCTGTTATCTTATGATCCGACAGTTCGCGAGTCCAGGATATGAATTTGCC 277
 QY 181 AGAAAGCGTCGGGAAATCTGCTGACAAACATTAACATCCAGTCCGAAATTAATGGA 240
 DB 278 AGAAAGCGTCGGGAAATCTGCTGACAAACATTAACATCCAGTCCGAAATTAATGGA 337
 QY 241 GTACTCCAGAAAGTATTAATCTGCTGATGAGAGCCGCGGCCCAAGAGCGTCCACA 300
 DB 338 GTACTCCAGAAAGTATTAATCTGCTGATGAGAGCCGCGGCCCAAGAGCGTCCACA 397
 QY 301 GGTGCGCCCTGTAAGATGCGCGCTGACGAACTTCACTTATGGAATCAACACGC 360
 DB 398 GGTGCGCCCTGTAAGATGCGCGCTGACGAACTTCACTTATGGAATCAACACGC 457
 QY 361 AAGACTATGACTCGTGAATCTCAAAATTAATGTTGAATGCGGAGAAACCATTTGAATG 420
 DB 458 AAGACTATGACTCGTGAATCTCAAAATTAATGTTGAATGCGGAGAAACCATTTGAATG 517
 QY 421 ATAGCTGACGCTTTTGAAGAAGATTAATCACTTGGTAGAGCATGGTTAAAGACACATG 480
 DB 518 ATAGCTGACGCTTTTGAAGAAGATTAATCACTTGGTAGAGCATGGTTAAAGACACATG 577
 QY 481 AATAATGAA 489
 DB 578 AATAATGAA 586

RESULT 8
 CBI02169 562 bp mRNA linear EST 28-JAN-2003
 LOCUS MW_50005@B3 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA sequence.
 DEFINITION

ACCESSION CB102169
 VERSION CB102169.1 GI:27927976
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans
 ORGANISM Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 562)
 Walhout,A.J., Sordella,R., Lu,X., Hartley,J.L., Temple,G.F.,
 Brasch,M.A., Thierry-Mieg,N. and Vidal,M.
 Protein interaction mapping in C. elegans using proteins involved
 in vulval development
 JOURNAL Science 287 (5450), 116-122 (2000)
 MEDLINE 20082953
 PUBMED 10615043
 COMMENT Contact: Vidal M
 Marc Vidal Laboratory
 Dana Farber Cancer Institute
 1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
 Tel: 617 632 5180
 Fax: 617 632 5739
 Email: Marc.Vidal@dfci.harvard.edu
 Trace dyp10315.x with Bait lin-53
 POLYA=No.
 FEATURES
 source Location/Qualifiers
 1..562
 /organism="Caenorhabditis elegans"
 /mol_type="mRNA"
 /strain="N2"
 /db_xref="taxon:6239"
 /sex="Hermaphrodite and male"
 /tissue_type="whole animal"
 /dev_stage="mixed stage"
 /clone_1ib="AD-wrmCDNA"
 /note="The AD-wrmCDNA library was generated with poly(A)+
 RNA isolated from both hermaphrodite and male N2 worms of
 all larval stages, embryos, adults and dauers and the
 subsequent generation of cDNAs by poly(A) priming. The
 cDNAs were cloned into pPC86"
 ORIGIN
 Query Match 50.6%; Score 483; DB 6; Length 562;
 Best Local Similarity 100.0%; Pred. No. 2,3e-116;
 Matches 483; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 CAGAAATAGATCCACTTGGCCGAGTTCTTGGTTCCAGAAACGAGATCGAAATGCTCTGC 64
 DB 80 CAGAAATAGATCCACTTGGCCGAGTTCTTGGTTCCAGAAACGAGATCGAAATGCTCTGC 139
 QY 65 AAAATATCATGATGAAGGAGGAGTCCACTTCATTGGAATGCCCAAGCAAGAAACTCA 124
 DB 140 AAAATATCATGATGAAGGAGGAGTCCACTTCATTGGAATGCCCAAGCAAGAAACTCA 199
 QY 125 CATCCCTGTATCTATGATCCGACAGTTCCGAGTCCCGATGATGAAATTCGCCAGAA 184
 DB 200 CATCCCTGTATCTATGATCCGACAGTTCCGAGTCCCGATGATGAAATTCGCCAGAA 259
 QY 185 AAGCTCTGGAATCTGCTGACCAACCTTAACAATCAATCCATGGAAATTAATGGAGTAC 244
 DB 260 AAGCTCTGGAATCTGCTGACCAACCTTAACAATCAATCCATGGAAATTAATGGAGTAC 319
 QY 245 TCCCAAGAAATTAATCTGCTGATGAGAAAGCCCGGCGCCCAAGACGTCCACAGGTC 304
 DB 320 TCCCAAGAAATTAATCTGCTGATGAGAAAGCCCGGCGCCCAAGACGTCCACAGGTC 379
 QY 305 GCCCTCTGAAGATCGCGGTCAAGATCTTCACTTAATGAGATCCACCGCAAGA 364
 DB 380 GCCCTCTGAAGATCGCGGTCAAGATCTTCACTTAATGAGATCCACCGCAAGA 439
 QY 365 CTATGATCTGATGATTAATTAATGTTGAATGCGTGGAAAAACCATGGAATGATAG 424
 DB 440 CTATGATCTGATGATTAATTAATGTTGAATGCGTGGAAAAACCATGGAATGATAG 499

QY 425 CTGACGTTTGAAGAGATATTCCTGTTAGAGCATGGGTAAAGACATGATAA 484
 DB 500 CTGACGTTTGAAGAGATATTCCTGTTAGAGCATGGGTAAAGACATGATAA 559
 QY 485 ATG 487
 DB 560 ATG 562
 RESULT 9
 LOCUS CB102055
 DEFINITION MWV_SQ001081 AD-wrmCDNA Caenorhabditis elegans CDNA, mRNA sequence.
 ACCESSION CB102055
 VERSION CB102055.1 GI:27927962
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans
 ORGANISM Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 495)
 Walhout,A.J., Sordella,R., Lu,X., Hartley,J.L., Temple,G.F.,
 Brasch,M.A., Thierry-Mieg,N. and Vidal,M.
 Protein interaction mapping in C. elegans using proteins involved
 in vulval development
 JOURNAL Science 287 (5450), 116-122 (2000)
 MEDLINE 20082953
 PUBMED 10615043
 COMMENT Contact: Vidal M
 Marc Vidal Laboratory
 Dana Farber Cancer Institute
 1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
 Tel: 617 632 5180
 Fax: 617 632 5739
 Email: Marc.Vidal@dfci.harvard.edu
 Trace lin14-AD with Bait lin-14
 POLYA=No.
 FEATURES
 source Location/Qualifiers
 1..495
 /organism="Caenorhabditis elegans"
 /mol_type="mRNA"
 /strain="N2"
 /db_xref="taxon:6239"
 /sex="Hermaphrodite and male"
 /tissue_type="whole animal"
 /dev_stage="mixed stage"
 /clone_1ib="AD-wrmCDNA"
 /note="The AD-wrmCDNA library was generated with poly(A)+
 RNA isolated from both hermaphrodite and male N2 worms of
 all larval stages, embryos, adults and dauers and the
 subsequent generation of cDNAs by poly(A) priming. The
 cDNAs were cloned into pPC86"
 ORIGIN
 Query Match 43.1%; Score 411; DB 6; Length 495;
 Best Local Similarity 100.0%; Pred. No. 2,3e-97;
 Matches 411; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGTCAAAATAGATCCACTTGGCCGAGTTCTTGGTTCCAGAAACGAGATCGAAATGCT 60
 DB 85 ATGTCAAAATAGATCCACTTGGCCGAGTTCTTGGTTCCAGAAACGAGATCGAAATGCT 144
 QY 61 CGTCAAAATAGATCCACTTGAAGCGGAGTCCACTTCATTGGAATGCCCAAGCAAGAAA 120
 DB 145 CGTCAAAATAGATCCACTTGAAGCGGAGTCCACTTCATTGGAATGCCCAAGCAAGAAA 204
 QY 121 CTCACATCCCTGTATCTATGATCCGACAGTTCCGAGTCCCGATGATGAAATTCGCC 180
 DB 205 CTCACATCCCTGTATCTATGATCCGACAGTTCCGAGTCCCGATGATGAAATTCGCC 264
 QY 181 AGAAAAAGTCTGGGAAATCTGCTGACCAACATTAACATCAATCCATCGGAAATTAATGGA 240
 DB 265 AGAAAAAGTCTGGGAAATCTGCTGACCAACATTAACATCAATCCATCGGAAATTAATGGA 324

Qy	241	GTACTCCAGAAAGATTATACTCGTGCTGATGAAAGCCCGCGCCGCAAGACGTCACCA	300
Db	325	GTACTCCAGAAAGATTATACTCGTGCTGATGAAAGCCCGCGCCGCAAGACGTCACCA	384
Qy	301	GGTGGCCCTCGTAAGATCCCGCGTCACAAATCTTGAACCTCATGTAAGAAATCACACGC	360
Db	385	GGTGGCCCTCGTAAGATCCCGCGTCACAAATCTTGAACCTCATGTAAGAAATCACACGC	444
Qy	361	AAGACTATGACTCGTGAATCTTAAATTAATGTTGAATTGCGTGGAAAACCA	411
Db	445	AAGACTATGACTCGTGAATCTTAAATTAATGTTGAATTGCGTGGAAAACCA	495
RESULT 10			
CB102059			
LOCUS	CB102059	485 bp	mRNA linear EST 28-JAN-2003
DEFINITION	MW S0001@B5 AD-wrmCDNA Caenorhabditis elegans cDNA, mRNA sequence.		
ACCESSION	CB102059		
VERSION	CB102059.1	GI:27927866	
KEYWORDS	EST..		
SOURCE	Caenorhabditis elegans		
ORGANISM	Caenorhabditis elegans		
REFERENCE	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitida; Rhabdilitidae; Peloderinae; Caenorhabditis.		
AUTHORS	1 (bases 1 to 485) Walhout,A.J., Sordella,R., Lu,X., Hattley,J.L., Temple,G.F., Brasch,M.A., Thierry-Mieg,N. and Vidal,M.		
TITLE	Protein interaction mapping in C. elegans using proteins involved in vulval development		
JOURNAL	Science 287 (5450), 116-122 (2000)		
MEDLINE	20082953		
PUBMED	10615043		
COMMENT	Contact: Vidal M Marc Vidal Laboratory Dana Farber Cancer Institute 1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA Tel: 617 632 5130 Fax: 617 632 5739 Email: Marc_Vidal@dfci.harvard.edu Trace Lin19-AD with Bait Lin19 POLYA=NO.		
FEATURES			
Source	Location/Qualifiers		
	1..485		
	/organism="Caenorhabditis elegans"		
	/mol_type="mRNA"		
	/strain="N2"		
	/db_xref="taxon:6239"		
	/sex="Hermaphrodite and male"		
	/tissue.type="whole animal"		
	/dev_stage="mixed stage"		
	/clone_lib="AD-wrmCDNA"		
	/note="The AD-wrmCDNA library was generated with poly(A)+ RNA isolated from both hermaphrodite and male N2 worms of all larval stages, embryos, adults and dauers and the subsequent generation of cDNAs by poly(A) priming. The cDNAs were cloned into pPC86"		
ORIGIN			
Query Match	41.6%; Score 397.2; DB 6; Length 485;		
Best Local Similarity	99.3%; Pred. No. 1e-93;		
Matches 399; Conservative 0; Mismatches 3; Indels 0; Gaps 0;			
Qy	1	ATGTCCAGAAATGATCCACTTCGCGAGTCTTGCTTCCAGAAAGCGAGATCGAAATGCT	60
Db	84	ATGTCCAGAAATGATCCACTTCGCGAGTCTTGCTTCCAGAAAGCGAGATCGAAATGCT	143
Qy	61	CGTCAAAATGATTCATTGATAGCGGAGGTCCACTTCCATTGATGATCGCCAAAGCAAAA	120
Db	144	COTCAAAATGATTCATTGATAGCGGAGGTCCACTTCCATTGATGATCGCCAAAGCAAAA	203
Qy	121	CTCACATCCCTGTTATCTTATGATTCGACAGTTCGCGAGTCACCGGATATGAAATTCGCC	180

Db	204	CTGACATCCCTGTTATCTCTATGATTCACGAGTCCGGAGTCAACGGATATGAAATTGGCC	265
Qy	181	AGAAAAGCTCTGGAAATCTGCTGACAAACCTAAACAATCAACCAATCGGAATATTGGA	240
Db	264	AGAAAAGCTCTGGGAAATCTGCTGACAAACCTAAACAATCAACCAATCGGAATATTGGA	323
Qy	241	GTACTCCAGAAAGTTATATCTCGTGGTGAAGAAAGCCGGGGCCCAAGACGTCCACCA	300
Db	324	GTACTCCAGAAAGTTATATCTCGTGGTGAAGAAAGCCGGGGCCCAAGACGTCCACCA	383
Qy	301	GGTGGCCCTCGTAAGATGCGCGTCACAGAAATCTTCAACTTCACTTATGAATCACCAGC	360
Db	384	GGTGGCCCTCGTAAGATGCGCGTCACAGAAATCTTCAACTTCACTTATGAATCACCAGC	443
Qy	361	AAGACTATGACTCGTGAATTTCTTAAATATATGTTGAATTGGCT	402
Db	444	AAGACTATGACTCGTGAATTTCTTAAATATATGTTGAATTGGCT	485
RESULT 11			
AV202470	377 bp	mRNA	linear
LOCUS			
AV202470			
DEFINITION	AV202470 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA		
ACCESSION	AV202470		
VERSION	AV202470.1	GI:5586241	
KEYWORDS	EST.		
SOURCE	Caenorhabditis elegans		
ORGANISM	Caenorhabditis elegans		
REFERENCE	Eukaryotic Metazoa; Nematoda; Chromodorea; Rhabditiida; Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.		
AUTHORS	1 (bases 1 to 377) Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Mitsuki, H., Nishikaki, A., Motochashi, T., Zeng, Q., Watanabe, H., Sugimoto, A., Sano, M., Miyata, A., Mitani, Y., Iida, K., Uesugi, H., Sugiyama, Y. and Nomoto, H.		
TITLE	Expressed genes in C. elegans		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Yuji Kohara Genome Biology Lab. National Institute of Genetics Yata 1111, Mishima, Shizuoka 411, Japan Tel: 81-559-81-6854 Fax: 81-559-81-6855 Email: ykohara@lab.nig.ac.jp.		
FEATURES	Location/Qualifiers		
Source	1..377		
	/organism="Caenorhabditis elegans"		
	/mol_type="mRNA"		
	/strain="CB1489 him-8(e1489)"		
	/db_xref="taxon:6239"		
	/clone="YK550h7"		
	/sex="hermaphrodite, male"		
	/tissue_type="whole animal"		
	/dev_stage="varied"		
	/clone_idb="Yuji Kohara unpublished cDNA"		
ORIGIN			
Query Match	38.3%;	Score 365;	DB 1; Length 377;
Best local Similarity	99.7%;	Pred. No. 3e-85;	
Matches 376;	Conservative 0;	Mismatches 0;	Indels 1; Gaps 1;
Qy	23	CCGAGTCTTGCTCTCCAGAAAGCGAGAATCGAAATGCTCGTCAAAATGATTCATTGATTA	82
Db	1	CCGAGTCTTGCTCTCCAGAAAGCGAGAATCGAAATGCTCGTCAAAATGATTCATTGATTA	60
Qy	83	GCGAGGTCCACTTTCATTGGAATGCGCAAGACG-AAAATCTACATCCCTGTTATCTCAT	141
Db	61	GCGAGGTCCACTTTCATTGGAATGCGCAAGCAAAAAAATCATCATCCCTGTTATCTCAT	120
Qy	142	GATCCGACAGTTCGGGAGTCAACGGATATGAAATTCGCCAGAAAAAGCTCTGGGAAATCTG	201
Db	121	GATCCGACAGTTCGGGAGTCAACGGATATGAAATTCGCCAGAAAAAGCTCTGGGAAATCTG	180

QY 202 CTGACAAACATTAACATCAACCCATCGAAATTAATGAGTACTCCCAAGAAATTAATCT 261
DB 181 CTGACAAACATTAACATCAACCCATCGAAATTAATGAGTACTCCCAAGAAATTAATCT 240
QY 262 CGTGTGATGAAGAGCCCGGCGCAAGAGAGCGTCCAGAGTCCGCTCGTAAGATGCCG 321
DB 241 CGTGTGATGAAGAGCCCGGCGCAAGAGAGCGTCCAGAGTCCGCTCGTAAGATGCCG 300
QY 322 CGTCAAGATCTTCACTTCACTTATGGAATCACCAGCAAGACTTATGATCTCGTATCT 381
DB 301 CGTCAAGATCTTCACTTCACTTATGGAATCACCAGCAAGACTTATGATCTCGTATCT 360
QY 382 AAAATTATGTTGAATT 398
DB 361 AAAATTATGTTGAATT 377

RESULT 12 360 bp mRNA linear EST 18-OCT-1999
D74918
LOCUS CELK090A5F Yui Kohara unpublished cDNA:Strain N2 hermaphrodite
DEFINITION embryo Caenorhabditis elegans cDNA clone yk90a5 5', mRNA sequence.
ACCESSION D74918 GI:1120702
VERSION D74918.1
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 360)
AUTHORS Kohara,Y., Mitsuki,H., Nishigaki,A., Mochizuki,T., Sugimoto,A. and Tabara,H.

TITLE Toward an expression map of the C.elegans genome
JOURNAL Unpublished (1994)
COMMENT Contact: Yui Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykoha@lab.nig.ac.jp.
Location/Qualifiers

FEATURES
source
1..360
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk90a5"
/sex="hermaphrodite"
/dev_stage="embryo"
/clone_1ib="Yui Kohara unpublished cDNA:Strain N2 hermaphrodite embryo"

ORIGIN

Query Match 37.1%; Score 354; DB 7; Length 360;
Best Local Similarity 100.0%; Pred. No. 2.4e-82;
Matches 354; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGTGCAAAATGATCCATTGCGGAGTTCTTGTCTCCAGAAAGCGAGATCGAAATGCT 60
DB 6 ATGTGCAAAATGATCCATTGCGGAGTTCTTGTCTCCAGAAAGCGAGATCGAAATGCT 65
QY 61 CGTCAAAATGATCCATTGAGGAGGATCCATTGATGGAATGCCAAGCAGAAA 120
DB 66 CGTCAAAATGATCCATTGAGGAGGATCCATTGATGGAATGCCAAGCAGAAA 125
QY 121 CTCACATCCCTGTATCTATGATCGACAGTTCCGAGTCAACCGATATGAATTCGCC 180
DB 126 CTCACATCCCTGTATCTATGATCGACAGTTCCGAGTCAACCGATATGAATTCGCC 185
QY 181 AGAAAAGTCTGGAAATCTGCTGACACCAATTAACATCAACCATCGAAATTAATGGA 240

DB 186 AGAAAAGTCTGGAAATCTGCTGACACCAATTAACATCAACCATCGAAATTAATGGA 245
QY 241 GTACTCCCAAGAAATTAATCTGCTGATGAAGAGCCCGGCGCAAGAGCGTCCACCA 300
DB 246 GTACTCCCAAGAAATTAATCTGCTGATGAAGAGCCCGGCGCAAGAGCGTCCACCA 305
QY 301 GGTGCGCCCTCGTAAGATGCGCGTCAAGATCTTCACTTATGGAATCA 354
DB 306 GGTGCGCCCTCGTAAGATGCGCGTCAAGATCTTCACTTATGGAATCA 359

RESULT 13 360 bp mRNA linear EST 18-OCT-1999
D75085
LOCUS CELK096B5F Yui Kohara unpublished cDNA:Strain N2 hermaphrodite
DEFINITION embryo Caenorhabditis elegans cDNA clone yk96b5 5', mRNA sequence.
ACCESSION D75085 GI:1120869
VERSION D75085.1
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 360)
AUTHORS Kohara,Y., Mitsuki,H., Nishigaki,A., Mochizuki,T., Sugimoto,A. and Tabara,H.
Toward an expression map of the C.elegans genome
JOURNAL Unpublished (1994)
COMMENT Contact: Yui Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykoha@lab.nig.ac.jp.
Location/Qualifiers

FEATURES
source
1..360
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk96b5"
/sex="hermaphrodite"
/dev_stage="embryo"
/clone_1ib="Yui Kohara unpublished cDNA:Strain N2 hermaphrodite embryo"

ORIGIN

Query Match 37.1%; Score 354; DB 7; Length 360;
Best Local Similarity 98.3%; Pred. No. 2.4e-82;
Matches 354; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 9 AATGATCCACTTGGCCGAGTTCTTGTCTCCAGAAAGCGAGATCGAAATGCTCGTCAAAA 68
DB 1 AATGATCCACTTGGCCGAGTTCTTGTCTCCAGAAAGCGAGATCGAAATGCTCGTCAAAA 60
QY 69 TGATCATTTGATTAACGAGGATCCATTGGAATGCCAAGCAAGAAATCAATC 128
DB 61 TGATCATTTGATTAACGAGGATCCATTGGAATGCCAAGCAAGAAATCAATC 120
QY 129 CCTGTATCTATGATCCGACAGTTCCGAGATCCCGGATATGAATTCGCGCAGAAAACG 188
DB 121 CCTGTATCTATGATCCGACAGTTCCGAGATCCCGGATATGAATTCGCGCAGAAAACG 180
QY 189 TCTGGAAATCTGCTGACCAACATTAACATCAACCATCGAAATTAATGGAATGTA 248
DB 181 TCTGGAAATCTGCTGACCAACATTAACATCAACCATCGAAATTAATGGAATGTA 240
QY 249 AGAAGATTATCTGCTGATGAAGAGCCCGGCGCAAGAGCGTCCACGAGTCCGCC 308
DB 241 AGAAGATTATCTGCTGATGAAGAGCCCGGCGCAAGAGCGTCCACGAGTCCGCC 300
QY 309 TCGTAAGATGCGCGTCAAGATCTTCACTTATGGAATCAACGCAAGACTAT 368

Db 301 TCGTAAGATGCGGATGACGATCTTCACTTATGGAATCACCACCAAGACTAT 360

RESULT 14
CB102066 438 bp mRNA linear EST 28-JAN-2003
LOCUS CB102066
DEFINITION MW50001@B12 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA
ACCESSION CB102066
VERSION CB102066.1 GI:27927873
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
AUTHORS Rhabditioidea; Rhabditidae; Pelodierinae; Caenorhabditis.
1 (bases 1 to 438)
Walhout,A.J., Sordella,R., Lu,X., Hartley,J.L., Temple,G.F.,
Brasch,M.A., Thierry-Mieg,N., and Vidal,M.
Protein interaction mapping in C. elegans using proteins involved
in vulval development
Science 287 (5450), 116-122 (2000)

JOURNAL
MEDLINE
PUBMED
20082953
10615043

COMMENT Contact: Vidal M
Marc Vidal Laboratory
Dana Farber Cancer Institute
1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
Tel: 617 632 5180
Fax: 617 632 5739
Email: Marc.Vidal@dfci.harvard.edu
Trace Lin8-AD with Bait Lin-8
POLYA=NO.

FEATURES
source
1..438
Location/Qualifiers
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/sex="Hermaphrodite and male"
/tissue_type="whole animal"
/dev_stage="mixed stage"
/clone_lib="AD-wrmcDNA"
/note="The AD-wrmcDNA library was generated with poly(A)+
RNA isolated from both hermaphrodite and male N2 worms of
all larval stages, embryos, adults and dauers and the
subsequent generation of cDNAs by poly(A) priming. The
cDNAs were cloned into pPC86"

ORIGIN
Query Match 37.1%; Score 354; DB 6; Length 438;
Best Local Similarity 100.0%; Pred. No. 2.5e-82;
Matches 354; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTCAAAATAGATTCACCTTCCGAGTTCTTCTTCAGAAAGCGAGATCGAAATGCT 60
DB 85 ATGTCAAAATAGATTCACCTTCCGAGTTCTTCTTCAGAAAGCGAGATCGAAATGCT 144
QY 61 CGTCAAAATGATTCATGATTAAGCGAGGTCCTTCATTCATGGAATGCCAAGAGAAA 120
DB 145 CGTCAAAATGATTCATGATTAAGCGAGGTCCTTCATTCATGGAATGCCAAGAGAAA 204
QY 121 CTCACATCCCTGTTATCTTATGATCCGACAGTTCCGAGTCAACCGATATGAATTTGCC 180
DB 205 CTCACATCCCTGTTATCTTATGATCCGACAGTTCCGAGTCAACCGATATGAATTTGCC 264
QY 181 AGAAAAGCTGTGGGAAATCTGCTGACAAACATTAACATCAACCATGGGAAATATTTGGA 240
DB 265 AGAAAAGCTGTGGGAAATCTGCTGACAAACATTAACATCAACCATGGGAAATATTTGGA 324
QY 241 GTACTCCAGAAAGATTATCTGCTGATGATGAAGCCCGGGGCCCAAGAGCGTCCACCA 300
DB 325 GTACTCCAGAAAGATTATCTGCTGATGATGAAGCCCGGGGCCCAAGAGCGTCCACCA 384

QY 301 GGTGCGCCTGTGAAGATGCGCGGTCAGGAATTTCACTTCACTTATGGAATCA 354
DB 385 GGTGCGCCTGTGAAGATGCGCGGTCAGGAATTTCACTTCACTTATGGAATCA 438

RESULT 15
CB102067 433 bp mRNA linear EST 28-JAN-2003
LOCUS CB102067
DEFINITION MW50001@C1 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA sequence.
ACCESSION CB102067
VERSION CB102067.1 GI:27927874
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
AUTHORS Rhabditioidea; Rhabditidae; Pelodierinae; Caenorhabditis.
1 (bases 1 to 433)
Walhout,A.J., Sordella,R., Lu,X., Hartley,J.L., Temple,G.F.,
Brasch,M.A., Thierry-Mieg,N., and Vidal,M.
Protein interaction mapping in C. elegans using proteins involved
in vulval development
Science 287 (5450), 116-122 (2000)

JOURNAL
MEDLINE
PUBMED
20082953
10615043

COMMENT Contact: Vidal M
Marc Vidal Laboratory
Dana Farber Cancer Institute
1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
Tel: 617 632 5180
Fax: 617 632 5739
Email: Marc.Vidal@dfci.harvard.edu
Trace Lin9-AD with Bait Lin-9
POLYA=NO.

FEATURES
source
1..433
Location/Qualifiers
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/sex="Hermaphrodite and male"
/tissue_type="whole animal"
/dev_stage="mixed stage"
/clone_lib="AD-wrmcDNA"
/note="The AD-wrmcDNA library was generated with poly(A)+
RNA isolated from both hermaphrodite and male N2 worms of
all larval stages, embryos, adults and dauers and the
subsequent generation of cDNAs by poly(A) priming. The
cDNAs were cloned into pPC86"

ORIGIN
Query Match 37.0%; Score 353; DB 6; Length 433;
Best Local Similarity 100.0%; Pred. No. 4.6e-82;
Matches 353; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTCAAAATAGATTCACCTTCCGAGTTCTTCTTCAGAAAGCGAGATCGAAATGCT 60
DB 81 ATGTCAAAATAGATTCACCTTCCGAGTTCTTCTTCAGAAAGCGAGATCGAAATGCT 140
QY 61 CGTCAAAATGATTCATGATTAAGCGAGGTCCTTCATTCATGGAATGCCAAGAGAAA 120
DB 141 CGTCAAAATGATTCATGATTAAGCGAGGTCCTTCATTCATGGAATGCCAAGAGAAA 200
QY 121 CTCACATCCCTGTTATCTTATGATCCGACAGTTCCGAGTCAACCGATATGAATTTGCC 180
DB 201 CTCACATCCCTGTTATCTTATGATCCGACAGTTCCGAGTCAACCGATATGAATTTGCC 260
QY 181 AGAAAAGCTGTGGGAAATCTGCTGACAAACATTAACATCAACCATGGGAAATATTTGGA 240
DB 261 AGAAAAGCTGTGGGAAATCTGCTGACAAACATTAACATCAACCATGGGAAATATTTGGA 320
QY 241 GTACTCCAGAAAGATTATCTGCTGATGATGAAGCCCGGGGCCCAAGAGCGTCCACCA 300

Db 321 GTACTCCAGAGATATATCTGTCGTGATGAGAGCCCGGCGCCAGAGAGCTCACCA 380
Qy 301 GGTGCCCCCTCGTAAGATGCCGCGTCACGATCTTCACTTATGGAATC 353
|||
Db 381 GGTGCCCCCTCGTAAGATGCCGCGTCACGATCTTCACTTATGGAATC 433
|||

Search completed: April 9, 2005, 06:18:14
Job time : 3734 secs

```

Query Match      100.0%; Score 954; DB 9; Length 954;
Best Local Similarity 100.0%; Prd. No. 9,4e-241;
Matches 954; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY      1  ATGTGAGAAATGATGCCACTTGGCCAGTCTTGTGCTTCAGAAAGAAGAGATCGAAATGCT 60
Db       1  ATGTGAGAAATGATGCCACTTGGCCAGTCTTGTGCTTCAGAAAGAAGAGATCGAAATGCT 60
QY      61  CGTCAAAATGATCCATTGATTAAGCGGAGGTCACTTCATTGGAATGCCCAAGCAAAA 120

```

Db 61 CGTCAAAATGATCCATTGATTAAGCGAGGTCCACTTCCATTGGAATGCCAAGAGAAA 120
OY 121 CTCACATCCCGTATATCTATATGATCCGACAGTCCGAGTCAACGGATATGAATTCGCC 180
Db 121 CTCACATCCCGTATATCTATATGATCCGACAGTCCGAGTCAACGGATATGAATTCGCC 180
OY 181 AGAAAACTCTGGGAAATCTGCTGACCAACATTAATAATCCATCGGAAATTAATTGGA 240
Db 181 AGAAAACTCTGGGAAATCTGCTGACCAACATTAATAATCCATCGGAAATTAATTGGA 240
OY 241 GATCTCCCAAGAGATTAATCTGCTGATGAAGAGCCGGGCGCAAGAGAGTCCACCA 300
Db 241 GATCTCCCAAGAGATTAATCTGCTGATGAAGAGCCGGGCGCAAGAGAGTCCACCA 300
OY 301 GGTGCGCCCTCTGAAGATCCGCGTACAGAACTTTCATCTTATGAAATCAACACGC 360
Db 301 GGTGCGCCCTCTGAAGATCCGCGTACAGAACTTTCATCTTATGAAATCAACACGC 360
OY 361 AAGACTATGACTGTGATTTCTAAATTAATGTTTGAATTCGTTGAATTCGAAATG 420
Db 361 AAGACTATGACTGTGATTTCTAAATTAATGTTTGAATTCGTTGAATTCGAAATG 420
OY 421 ATAGCTGACGTTTGAAGAGATTAATCTGTTAGAGCATGAGGTTAAAGACACATG 480
Db 421 ATAGCTGACGTTTGAAGAGATTAATCTGTTAGAGCATGAGGTTAAAGACACATG 480
OY 481 AATAATGATATGAACCAATTAAGCTCAAGAGACATATGACCGAATCTGCTGTT 540
Db 481 AATAATGATATGAACCAATTAAGCTCAAGAGACATATGACCGAATCTGCTGTT 540
OY 541 GATTAATCTTGATGATGCGGAGATTCATGCAATGCCAGTCCGATTAATCTCTGAG 600
Db 541 GATTAATCTTGATGATGCGGAGATTCATGCAATGCCAGTCCGATTAATCTCTGAG 600
OY 601 GTATTAATCTTGATGATGCGGAGATTCATGCAATGCCAGTCCGATTAATCTCTGAG 660
Db 601 GTATTAATCTTGATGATGCGGAGATTCATGCAATGCCAGTCCGATTAATCTCTGAG 660
OY 661 ACAGATTTGAAAAATGATTAATCAATTCGTCATTTGGAACCAAGTGGTGTGCT 720
Db 661 ACAGATTTGAAAAATGATTAATCAATTCGTCATTTGGAACCAAGTGGTGTGCT 720
OY 721 CATCAACGTCGTCGAGTCTGCCATGCAAGAGATGAGATTAATCAACAAATCTAC 780
Db 721 CATCAACGTCGTCGAGTCTGCCATGCAAGAGATGAGATTAATCAACAAATCTAC 780
OY 781 CAGCCTGAGAGTCTGAAAACTGTGAGCAAGCACTTGTCTTAATTAATTTCTAAT 840
Db 781 CAGCCTGAGAGTCTGAAAACTGTGAGCAAGCACTTGTCTTAATTAATTTCTAAT 840
OY 841 GTATTAATCTTAATCTCTTACTTCTGATCTTGTGATCAATGCTCTTAATTTCAAAA 900
Db 841 GTATTAATCTTAATCTCTTACTTCTGATCTTGTGATCAATGCTCTTAATTTCAAAA 900
OY 901 TCTCACTTTAAATTCATTAATTAATGAGTTTATTCATTAATCATCTTGTAC 954
Db 901 TCTCACTTTAAATTCATTAATTAATGAGTTTATTCATTAATCATCTTGTAC 954

RESULT 2
US-10-311-455-2011/c
; Sequence 2011, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detecting
; FILE REFERENCE: 5013.1014
; CURRENT FILING DATE: US/10/311.455
; PRIOR APPLICATION NUMBER: PCT/EP01/07537

PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 2011
LENGTH: 6291
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2011

Query Match
Best Local Similarity 58.8%; Score 44.6; DB 15; Length 6292;
Matches 77; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

OY 812 CACTTGCTTATTAATTAATTTCTAAACATGTAATTTCAATTTATCTTACTTCTGATC 871
Db 3386 CATTATTAATTAATTAATTAATTTCTAAACATGTAATTTCAATTTATCTTACTTCTGATC 871
OY 872 TTGCTATCAGATGCTCTTATTTCAAAATTCACCTTAAATTCATTAATTAATGAGT 931
Db 3326 ATCCCATTAATTAATTTCTAATTTCTCAAAAAATTTTAACTCTATTAATTAACAC 3267
OY 932 TTATTCAAATA 942
Db 3266 TTTTAAACCA 3256

RESULT 3
US-10-221-714A-521/c
; Sequence 521, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; FILE REFERENCE: 5013.1005
; CURRENT FILING DATE: US/10/221.714A
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 521
LENGTH: 6292
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-521

Query Match
Best Local Similarity 4.7%; Score 44.6; DB 17; Length 6292;
Matches 77; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

OY 812 CACTTGCTTATTAATTAATTTCTAAACATGTAATTTCAATTTATCTTACTTCTGATC 871
Db 3386 CATTATTAATTAATTAATTTCTAAACATGTAATTTCAATTTATCTTACTTCTGATC 871

Qy	872	TTGCGATCAGCAGTCTTATTTCAAAATCGCACTTAAATCATTTAAATATATGGGT	931
Db	3326	ATCCCATAAATATCTATTCTCAAAAATTTTATTAAACCTCTATTATATAAACAAC	326
Qy	932	TTATTTCAATA	942
Db	3266	TTTTTAAACA	3256

RESULT 4
US-10-240-453-201

	Query Match	4.7%	Score 44.4;	DB 15;	Length 8346;
	Best Local Similarity	59.5%;	Pred. No. 1.8;		
	Matches	75;	Conservative	0;	Mismatches 51; Indels 0; Gaps 0;
Qy	817	GGTCTTAATTTAAATTCATACGATGATAATTCAATTTACTTCTAGACTTGCT	876		
Dd	3653	GGTTAATTTTTTAANGTTAATTTTGTTCGTTTTTTTTTTGTTTGAATTTGGGA	3712		
Qy	877	ATCACATGCTCTTAATTCAAAATCTCAGCTTAAATCATATAATATGGGTTAAT	936		
Dd	3713	TTAGATTTTTTTTTTTTTAAAAATTTTATTTTAGTTTGAGATAGTTTGTGA	3772		
Qy	937	CAAATA 942			
Dd	3773	TAAATA 3778			

RESULT 5
US-10-706-635--23
Sequence 23, Application US/10706635
Publication No. US20050014263A1
GENERAL INFORMATION:
APPLICANT: Moyer, Richard W.
APPLICANT: Li, Yi
TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
TITLE OF INVENTION: Vertebrate Cells
FILE REFERENCE: UF-221C1XCZ1
CURRENT APPLICATION NUMBER: US/10/706,635
CURRENT FILING DATE: 2003-11-12
PRIOR APPLICATION NUMBER: 09/086,651

```

; PRIOR FILING DATE: 1998-05-29
; PRIOR APPLICATION NUMBER: 09/662,254
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/224,479
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 50000
; TYPE: DNA
; ORGANISM: Amapacta moorei entomopoxvirus
US-10-706-635-23

```

Query Match	4.5%;	Score 42.6;	DB 19;	Length 50000;
Best Local Similarity	61.1%;	Pred. No. 14;		
Matches 69; Conservative	0;	Mismatches 44;	Indels 0;	Gaps 0;

QY 815 TTGGTCTATTAAATATCTAACGATGTAATTCATTATACCTGACTTCGTGAACTGG 874

Db 35861 TTGGTGGTCTCTAAATAATTAAATGTGATTTAAATTATTAATGGCGCATATATATATTT 36020

QY 875 CTATCAGATGTCTCTATTTCAAAAATCTCAGCTTTAAATTCATATTAATAT 927

Db 36021 TGAATCTAAATTCATTTCCAAAAATTTTCATTATTAATAATATATATAAATAAT 36073

```

RESULT 6
US-10-027-632-256799/C
; Sequence 256799, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218, 006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198, 676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 256799
; LENGTH: 3242
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-256799

```

Query Match	Best Local Similarity	4.4%;	Score 42.4;	DB 13;	Length 3242;
Matches	82; Conservative	0;	Mismatches	66;	Indels 0; Gaps 0;
QY	806 AGCAGCAGCTGGCTCTATTAAATATCTTAACATGATTAATTCATTTACTT	865			
DB	1943 AGCAGACAGAGCTCTGTTAAAGTTAAATATACATATTATTAATGTTGGCTTGCT	1884			
QY	866 CTGATCTGTATACACATGCTCTTATTTCAAAATCTACTTTAAATTCATATAATA	925			
DB	1883 AAATTTATGATACAGAGAGATTTTATTTATGTATAGATTTCAAAAGCATATAATG	1824			
QY	926 ATGGGTTATTCAAATACATCATCTTGA	953			

Db 1823 ATTATTTAATTAATCCCAATCATCTTCA 1796

RESULT 7

US-10-027-632-256800/c
Sequence 256800, Application US/10027632
Publication No. US2002019871A1

GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
Polymorphisms in the Human Genome

FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

PRIOR FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 256800

LENGTH: 3242

TYPE: DNA

ORGANISM: Human

US-10-027-632-256800

Query Match

Best Local Similarity 4.4%; Score 42.4; DB 13; Length 3242;
Pred. No. 3.6;

Matches 82; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 806 AGCAAGCAGCTTGCTTTAATTAATCTAATGTAATTCATTTATCTCTTACTT 865

DB 1943 AGCAAGCAGCAAGCTCTTGTAAAGTTTAAATACATATTAAATGTTGCTTGCT 1884

QY 866 CTGATCTTGCTATCAATGCTCTTATTCATTAATCTCAAAAATCTCACTTAAATTCATATAATA 925

DB 1883 AAAATTATGATACAGAAAGATTATTTATTTGATATGATATTCATTAAGACATATAATG 1824

QY 926 ATGGGTTTATTCATTAATCATCTTGA 953

DB 1823 ATTATTTAATTAATCCCAATCATCTTCA 1796

QY 806 AGCAAGCAGCTTGCTTTAATTAATCTAATGTAATTCATTTATCTCTTACTT 865

DB 1943 AGCAAGCAGCAAGCTCTTGTAAAGTTTAAATACATATTAAATGTTGCTTGCT 1884

QY 866 CTGATCTTGCTATCAATGCTCTTATTCATTAATCTCAAAAATCTCACTTAAATTCATATAATA 925

DB 1883 AAAATTATGATACAGAAAGATTATTTATTTGATATGATATTCATTAAGACATATAATG 1824

QY 926 ATGGGTTTATTCATTAATCATCTTGA 953

DB 1823 ATTATTTAATTAATCCCAATCATCTTCA 1796

QY 806 AGCAAGCAGCTTGCTTTAATTAATCTAATGTAATTCATTTATCTCTTACTT 865

DB 1943 AGCAAGCAGCAAGCTCTTGTAAAGTTTAAATACATATTAAATGTTGCTTGCT 1884

QY 866 CTGATCTTGCTATCAATGCTCTTATTCATTAATCTCAAAAATCTCACTTAAATTCATATAATA 925

DB 1883 AAAATTATGATACAGAAAGATTATTTATTTGATATGATATTCATTAAGACATATAATG 1824

QY 926 ATGGGTTTATTCATTAATCATCTTGA 953

DB 1823 ATTATTTAATTAATCCCAATCATCTTCA 1796

QY 806 AGCAAGCAGCTTGCTTTAATTAATCTAATGTAATTCATTTATCTCTTACTT 865

DB 1943 AGCAAGCAGCAAGCTCTTGTAAAGTTTAAATACATATTAAATGTTGCTTGCT 1884

QY 866 CTGATCTTGCTATCAATGCTCTTATTCATTAATCTCAAAAATCTCACTTAAATTCATATAATA 925

DB 1883 AAAATTATGATACAGAAAGATTATTTATTTGATATGATATTCATTAAGACATATAATG 1824

QY 926 ATGGGTTTATTCATTAATCATCTTGA 953

DB 1823 ATTATTTAATTAATCCCAATCATCTTCA 1796

QY 806 AGCAAGCAGCTTGCTTTAATTAATCTAATGTAATTCATTTATCTCTTACTT 865

DB 1943 AGCAAGCAGCAAGCTCTTGTAAAGTTTAAATACATATTAAATGTTGCTTGCT 1884

QY 866 CTGATCTTGCTATCAATGCTCTTATTCATTAATCTCAAAAATCTCACTTAAATTCATATAATA 925

DB 1883 AAAATTATGATACAGAAAGATTATTTATTTGATATGATATTCATTAAGACATATAATG 1824

QY 926 ATGGGTTTATTCATTAATCATCTTGA 953

DB 1823 ATTATTTAATTAATCCCAATCATCTTCA 1796

QY 806 AGCAAGCAGCTTGCTTTAATTAATCTAATGTAATTCATTTATCTCTTACTT 865

DB 1943 AGCAAGCAGCAAGCTCTTGTAAAGTTTAAATACATATTAAATGTTGCTTGCT 1884

QY 866 CTGATCTTGCTATCAATGCTCTTATTCATTAATCTCAAAAATCTCACTTAAATTCATATAATA 925

DB 1883 AAAATTATGATACAGAAAGATTATTTATTTGATATGATATTCATTAAGACATATAATG 1824

QY 926 ATGGGTTTATTCATTAATCATCTTGA 953

DB 1823 ATTATTTAATTAATCCCAATCATCTTCA 1796

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 256799

LENGTH: 3242

TYPE: DNA

ORGANISM: Human

US-10-027-632-256799

Query Match

Best Local Similarity 4.4%; Score 42.4; DB 17; Length 3242;
Pred. No. 3.6;

Matches 82; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 806 AGCAAGCAGCTTGCTTTAATTAATCTAATGTAATTCATTTATCTCTTACTT 865

DB 1943 AGCAAGCAGCAAGCTCTTGTAAAGTTTAAATACATATTAAATGTTGCTTGCT 1884

QY 866 CTGATCTTGCTATCAATGCTCTTATTCATTAATCTCAAAAATCTCACTTAAATTCATATAATA 925

DB 1883 AAAATTATGATACAGAAAGATTATTTATTTGATATGATATTCATTAAGACATATAATG 1824

QY 926 ATGGGTTTATTCATTAATCATCTTGA 953

DB 1823 ATTATTTAATTAATCCCAATCATCTTCA 1796

QY 806 AGCAAGCAGCTTGCTTTAATTAATCTAATGTAATTCATTTATCTCTTACTT 865

DB 1943 AGCAAGCAGCAAGCTCTTGTAAAGTTTAAATACATATTAAATGTTGCTTGCT 1884

QY 866 CTGATCTTGCTATCAATGCTCTTATTCATTAATCTCAAAAATCTCACTTAAATTCATATAATA 925

DB 1883 AAAATTATGATACAGAAAGATTATTTATTTGATATGATATTCATTAAGACATATAATG 1824

QY 926 ATGGGTTTATTCATTAATCATCTTGA 953

DB 1823 ATTATTTAATTAATCCCAATCATCTTCA 1796

QY 806 AGCAAGCAGCTTGCTTTAATTAATCTAATGTAATTCATTTATCTCTTACTT 865

DB 1943 AGCAAGCAGCAAGCTCTTGTAAAGTTTAAATACATATTAAATGTTGCTTGCT 1884

QY 866 CTGATCTTGCTATCAATGCTCTTATTCATTAATCTCAAAAATCTCACTTAAATTCATATAATA 925

DB 1883 AAAATTATGATACAGAAAGATTATTTATTTGATATGATATTCATTAAGACATATAATG 1824

QY 926 ATGGGTTTATTCATTAATCATCTTGA 953

DB 1823 ATTATTTAATTAATCCCAATCATCTTCA 1796

QY 806 AGCAAGCAGCTTGCTTTAATTAATCTAATGTAATTCATTTATCTCTTACTT 865

DB 1943 AGCAAGCAGCAAGCTCTTGTAAAGTTTAAATACATATTAAATGTTGCTTGCT 1884

QY 866 CTGATCTTGCTATCAATGCTCTTATTCATTAATCTCAAAAATCTCACTTAAATTCATATAATA 925

DB 1883 AAAATTATGATACAGAAAGATTATTTATTTGATATGATATTCATTAAGACATATAATG 1824

QY 926 ATGGGTTTATTCATTAATCATCTTGA 953

DB 1823 ATTATTTAATTAATCCCAATCATCTTCA 1796

QY 806 AGCAAGCAGCTTGCTTTAATTAATCTAATGTAATTCATTTATCTCTTACTT 865

DB 1943 AGCAAGCAGCAAGCTCTTGTAAAGTTTAAATACATATTAAATGTTGCTTGCT 1884

QY 866 CTGATCTTGCTATCAATGCTCTTATTCATTAATCTCAAAAATCTCACTTAAATTCATATAATA 925

DB 1883 AAAATTATGATACAGAAAGATTATTTATTTGATATGATATTCATTAAGACATATAATG 1824

QY 926 ATGGGTTTATTCATTAATCATCTTGA 953

DB 1823 ATTATTTAATTAATCCCAATCATCTTCA 1796

QY 806 AGCAAGCAGCTTGCTTTAATTAATCTAATGTAATTCATTTATCTCTTACTT 865

DB 1943 AGCAAGCAGCAAGCTCTTGTAAAGTTTAAATACATATTAAATGTTGCTTGCT 1884

QY 866 CTGATCTTGCTATCAATGCTCTTATTCATTAATCTCAAAAATCTCACTTAAATTCATATAATA 925

DB 1883 AAAATTATGATACAGAAAGATTATTTATTTGATATGATATTCATTAAGACATATAATG 1824

QY 926 ATGGGTTTATTCATTAATCATCTTGA 953

DB 1823 ATTATTTAATTAATCCCAATCATCTTCA 1796

QY 806 AGCAAGCAGCTTGCTTTAATTAATCTAATGTAATTCATTTATCTCTTACTT 865

DB 1943 AGCAAGCAGCAAGCTCTTGTAAAGTTTAAATACATATTAAATGTTGCTTGCT 1884

QY 866 CTGATCTTGCTATCAATGCTCTTATTCATTAATCTCAAAAATCTCACTTAAATTCATATAATA 925

DB 1883 AAAATTATGATACAGAAAGATTATTTATTTGATATGATATTCATTAAGACATATAATG 1824

QY 926 ATGGGTTTATTCATTAATCATCTTGA 953

DB 1823 ATTATTTAATTAATCCCAATCATCTTCA 1796

QY 806 AGCAAGCAGCTTGCTTTAATTAATCTAATGTAATTCATTTATCTCTTACTT 865

DB 1943 AGCAAGCAGCAAGCTCTTGTAAAGTTTAAATACATATTAAATGTTGCTTGCT 1884

QY 866 CTGATCTTGCTATCAATGCTCTTATTCATTAATCTCAAAAATCTCACTTAAATTCATATAATA 925

DB 1883 AAAATTATGATACAGAAAGATTATTTATTTGATATGATATTCATTAAGACATATAATG 1824

QY 926 ATGGGTTTATTCATTAATCATCTTGA 953

DB 1823 ATTATTTAATTAATCCCAATCATCTTCA 1796

QY 806 AGCAAGCAGCTTGCTTTAATTAATCTAATGTAATTCATTTATCTCTTACTT 865

DB 1943 AGCAAGCAGCAAGCTCTTGTAAAGTTTAAATACATATTAAATGTTGCTTGCT 1884

QY 866 CTGATCTTGCTATCAATGCTCTTATTCATTAATCTCAAAAATCTCACTTAAATTCATATAATA 925

DB 1883 AAAATTATGATACAGAAAGATTATTTATTTGATATGATATTCATTAAGACATATAATG 1824

QY 926 ATGGGTTTATTCATTAATCATCTTGA 953

DB 1823 ATTATTTAATTAATCCCAATCATCTTCA 1796

QY 806 AGCAAGCAGCTTGCTTTAATTAATCTAATGTAATTCATTTATCTCTTACTT 865

DB 1943 AGCAAGCAGCAAGCTCTTGTAAAGTTTAAATACATATTAAATGTTGCTTGCT 1884

QY 866 CTGATCTTGCTATCAATGCTCTTATTCATTAATCTCAAAAATCTCACTTAAATTCATATAATA 925

DB 1883 AAAATTATGATACAGAAAGATTATTTATTTGATATGATATTCATTAAGACATATAATG 1824

QY 926 ATGGGTTTATTCATTAATCATCTTGA 953

Db 1883 AAAATTGATGACAGAGATTTTATTGATAGATATTCAGAAACATATAATG 1824
Qy 926 ATGGTTTATTCAGATCATCTTGA 953
Db 1823 ATTATCTTAATYACCCAAATCATCTTCA 1796

RESULT 10
US-10-239-676-147/C
; Sequence 147, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239, 676
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 147
; LENGTH: 6874
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-147

Query Match 4.4%; Score 42.4; DB 14; Length 6874;
Best Local Similarity 50.5%; Pred. No. 5.4;
Matches 103; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

Qy 747 TGCAGAGCATGACATTAATCAACAATCTACAGCTGGAGAGTCGAAACTGCGA 806
Db 2842 TACATTTATATTAATTAATTAACCCCTACAAAATTACCATTTTATTAATCAAAAAATTAAT 2783
Qy 807 GCAAGCATTGCTTATTTAAATATCTAACATGTAATTTCAATTTATCTTACTTTC 866
Db 2782 ATTATCAATTTTCATATTAATCAATATATATATTAATTAACAATTTAACTACATTT 2723
Qy 867 TGATCTTGCTATCACAGTCTCTTATTTCAAAAATCTCACTTTAAATTCATATTAATA 926
Db 2722 CTTACTTAATATCAATCCAAATCTTAATAAATACTTAATAAATATATATCAAAATTA 2663
Qy 927 TGGGTTTATCAATACATCATCT 950
Db 2662 TCAATAAATATATATATATTTT 2639

RESULT 11
US-10-240-453-163/C
; Sequence 163, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
; TITLE OF INVENTION: Transcription
; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated
; FILE REFERENCE: 5013.1009
; OTHER INFORMATION: with DNA Transcription
; CURRENT APPLICATION NUMBER: US/10/240, 453

; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 163
; LENGTH: 6874
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-163

Query Match 4.4%; Score 42.4; DB 15; Length 6874;
Best Local Similarity 50.5%; Pred. No. 5.4;
Matches 103; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

Qy 747 TGCAGAGCATGACATTAATCAACAATCTACAGCTGGAGAGTCGAAACTGCGA 806
Db 2842 TACATTTATATTAATTAATTAACCCCTACAAAATTACCATTTTATTAATCAAAAAATTAAT 2783
Qy 807 GCAAGCATTGCTTATTTAAATATCTAACATGTAATTTCAATTTATCTTACTTTC 866
Db 2782 ATTATCAATTTTCATATTAATCAATATATATATTAATTAACAATTTAACTACATTT 2723
Qy 867 TGATCTTGCTATCACAGTCTCTTATTTCAAAAATCTCACTTTAAATTCATATTAATA 926
Db 2722 CTTACTTAATATCAATCCAAATCTTAATAAATACTTAATAAATATATATCAAAATTA 2663
Qy 927 TGGGTTTATCAATACATCATCT 950
Db 2662 TCAATAAATATATATATATTTT 2639

RESULT 12
US-10-221-613-295/C
; Sequence 295, Application US/10221613
; Publication No. US20040029123A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
; FILE REFERENCE: 5013.1004
; CURRENT APPLICATION NUMBER: US/10/221, 613
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: PCT/EP01/02945
; DE 10013847.00
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-03-15
; 2000-03-15
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 428
; SEQ ID NO 295
; LENGTH: 6874
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-613-295

Query Match 4.4%; Score 42.4; DB 17; Length 6874;
Best Local Similarity 50.5%; Pred. No. 5.4; Indels 0; Gaps 0;
Matches 103; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 747 TGCAGAGCATTGATTAATCAACAAAATCTACGAGCTGGAGTCCAAAATCTGCA 806
DB 2842 TACATTAATTAATTAACCTTACCAAAATTAATCAATTTTAAATCAAAAATTAAT 2783
QY 807 GCAGAGCTGGCTTATTAATTAATCAACATTAATTTCAATTAATCTTACTTC 866
DB 2782 ATTATCAATTTTCAAAATTCATATATATATATTAATTAACATATTAATCACTATT 2723
QY 867 TGAATCTTGCATCACTGCTCTTATTTCAAAAATCTCACTTTAAATCATATAATA 926
DB 2722 CTACATATAATCAATCAATCCAAAATCTTAAATAATTAATTAATTAACAAATA 2663
QY 927 TGGGTTTATTCAAATATCATCT 950
DB 2662 TCAATTAATTAATTAATATATTTT 2639

RESULT 13

US-10-311-455-1844/C
; Sequence 1844, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1844
; LENGTH: 5511
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1844

Query Match 4.3%; Score 41; DB 15; Length 5511;
Best Local Similarity 58.7%; Pred. No. 11; Indels 0; Gaps 0;
Matches 71; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 824 TTTAAATATCTAATGTAATTTCAATTTATCTTCACTTTCGATCTGATCAATCAT 883
DB 4483 TTAATTAATTAATTAATTTTAAATTTTAAATTTTCCATTCATATCTTAT 4424
QY 884 GTCTCTTATTTCAAAAATCTCACTTAATTAATCATATAAAGGGTTTATTCAAATAC 943
DB 4423 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTAAAC 4364
QY 944 A 944
DB 4363 A 4363

RESULT 14

US-10-294-934-404/C
; Sequence 404, Application US/10294934
; Publication No. US20040038231A1
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta

APPLICANT: Chumakov, Ilya
APPLICANT: Bougueleret, Lydie
APPLICANT: Cohen, Amick

FILE REFERENCE: 62. US4.DIV
CURRENT APPLICATION NUMBER: US/10/294,934
CURRENT FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US 09/671,317
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US 09/536,178
PRIOR FILING DATE: 2000-03-23
PRIOR APPLICATION NUMBER: PCT/IB00/00403
PRIOR FILING DATE: 2000-03-24
PRIOR APPLICATION NUMBER: US 60/126,269
PRIOR FILING DATE: 1999-03-25
PRIOR APPLICATION NUMBER: US 60/131,961
PRIOR FILING DATE: 1999-04-30
NUMBER OF SEQ ID NOS: 977
SOFTWARE: Patent.pm
SEQ ID NO 404
LENGTH: 1001
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 503
OTHER INFORMATION: 10-453-330 : polymorphic base C or T
FEATURE:
NAME/KEY: misc_binding
LOCATION: 483..502
OTHER INFORMATION: 10-453-330. misc1, potential
FEATURE:
NAME/KEY: misc_binding
LOCATION: 504..523
OTHER INFORMATION: 10-453-330. misc2, potential complement
FEATURE:
NAME/KEY: primer_bind
LOCATION: 172..189
OTHER INFORMATION: upstream amplification primer
FEATURE:
NAME/KEY: primer_bind
LOCATION: 578..597
OTHER INFORMATION: downstream amplification primer, complement
FEATURE:
NAME/KEY: misc_binding
LOCATION: 491..515
OTHER INFORMATION: 10-453-330 potential probe
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19..31,101,212,337,520,661
OTHER INFORMATION: n=a, g, c or t
US-10-294-934-404

Query Match 4.2%; Score 40.2; DB 17; Length 1001;
Best Local Similarity 60.6%; Pred. No. 7.3; Indels 0; Gaps 0;
Matches 66; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 819 TCTTATTAATATCTTAACATGTAATTTCAATTTATCTTACTTTCGATCTGCTAT 878
DB 802 TTTAATTAACAACTTTATTAATTTTCAAGATTAATCTCTTCTACATAGGATTC 743
QY 879 CACATGCTCTTATTTCAAAAATCTCACTTAATTAATTAATTAATTAAT 927
DB 742 ATTATGTCAAATTTTAAAGAAATGTAATTTTATATTAATTAATCAATCAT 694

RESULT 15

US-10-311-455-1196/C
; Sequence 1196, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian

APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect
FILE REFERENCE: 5013.1014
CURRENT FILING DATE: 2002-12-16
PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 1196
LENGTH: 7560
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1196

Query Match 4.2%; Score 40.2; DB 15; Length 7560;
Best Local Similarity 55.3%; Pred. No. 22;
Matches 78; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Oy 813 ACTGGCTTATTAAATATCTAATGATGATTTCAATTTCAATTTCTACTTCTGATCT 872
Db 2401 ACATATTATATATATAATAATATTATCTTTCTTTTCCATTTTTCATATATCTTTT 2342
Oy 873 TGCTATCACATGCTCTTATTTCAAAAATCTCACTTAAATTCATATATAATGGGTT 932
Db 2341 TAAATTATAAATTATTATTTCAAAATCTAAATTTATAATAACATAATCACTCAT 2282
Oy 933 TATCAATACATCATCTTGA 953
Db 2281 TATTAAAAATATATCAAAATTAA 2261

Search completed: April 9, 2005, 07:53:22
Job time : 672 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 9, 2005, 02:28:09 ; Search time 202 Seconds
(without alignments)
7727.762 Million cell updates/sec

Title: US-09-087-136-2

Perfect score: 954
Sequence: 1 atgtcagaatagatcaccac.....ttcaatcacatcattgac 954

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCURS.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51.2	5.4	1141	US-09-806-708B-22	Sequence 22, Appl
2	40.8	4.3	38371	US-09-949-016-12061	Sequence 12061, A
3	40.8	4.3	38371	US-09-949-016-12488	Sequence 12488, A
4	40.8	4.3	38371	US-09-949-016-15596	Sequence 15596, A
5	40.8	4.3	38371	US-09-949-016-15597	Sequence 15597, A
6	40.4	4.2	601	US-09-949-016-94473	Sequence 94473, A
7	40.4	4.2	601	US-09-949-016-94474	Sequence 94474, A
8	40.4	4.2	20441	US-09-949-016-14476	Sequence 14476, A
9	40.2	4.2	1001	US-09-671-317-404	Sequence 404, App
10	40	4.2	15236	US-09-949-016-16778	Sequence 16778, A
11	40	4.2	640681	US-09-790-988-1	Sequence 1, Appl1
12	39.6	4.2	192700	US-09-949-016-11820	Sequence 11820, A
13	39.6	4.2	192704	US-09-949-016-17182	Sequence 17182, A
14	38.6	4.0	33044	US-09-949-016-17613	Sequence 17613, A
15	38.6	4.0	69813	US-09-949-016-12455	Sequence 12455, A
16	38.6	4.0	69813	US-09-949-016-13905	Sequence 13905, A
17	38.6	4.0	69813	US-09-949-016-13906	Sequence 13906, A
18	38.6	4.0	69813	US-09-949-016-12861	Sequence 12861, A
19	38.2	4.0	601	US-09-949-016-31836	Sequence 31836, A
20	38.2	4.0	601	US-09-949-016-40021	Sequence 40021, A
21	38.2	4.0	601	US-09-949-016-74003	Sequence 74003, A
22	38.2	4.0	601	US-09-949-016-74109	Sequence 74109, A
23	38.2	4.0	601	US-09-949-016-12853	Sequence 12853, A
24	38.2	4.0	20935	US-09-949-016-15383	Sequence 15383, A
25	37.2	3.9	1608	US-09-816-028A-18	Sequence 18, Appl
26	37.2	3.9	1608	US-09-816-028A-24	Sequence 24, Appl
27	37.2	3.9	1608	US-10-303-162-18	Sequence 18, Appl

C	28	37.2	3.9	1608	4	US-10-303-162-24	Sequence 24, Appl
C	29	37.2	3.9	1608	4	US-10-303-134-18	Sequence 18, Appl
C	30	37.2	3.9	1608	4	US-10-303-134-24	Sequence 24, Appl
C	31	37.2	3.9	18373	4	US-09-949-016-14338	Sequence 14338, A
C	32	37.2	3.9	18452	4	US-09-949-016-14337	Sequence 14337, A
C	33	37	3.9	601	4	US-09-949-016-198191	Sequence 198191, A
C	34	37	3.9	31407	4	US-09-949-016-17359	Sequence 17359, A
C	35	36.8	3.9	1664976	4	US-08-916-421B-1	Sequence 1, Appl1
C	36	36.8	3.9	1664976	4	US-08-916-421B-1	Sequence 1, Appl1
C	37	36.8	3.9	1664976	4	US-09-692-570-1	Sequence 1, Appl1
C	38	36.8	3.9	1664976	4	US-09-692-570-1	Sequence 1, Appl1
C	39	36.6	3.8	5300	1	US-08-766-014-1	Sequence 1, Appl1
C	40	36.6	3.8	227750	4	US-09-949-016-17175	Sequence 17175, A
C	41	36.2	3.8	225127	4	US-09-949-016-16480	Sequence 16480, A
C	42	36	3.8	601	4	US-09-949-016-180625	Sequence 180625, A
C	43	36	3.8	2287	4	US-09-023-655-105	Sequence 105, App
C	44	36	3.8	21723	4	US-09-949-016-16383	Sequence 16383, A
C	45	36	3.8	36618	4	US-09-949-016-16935	Sequence 16935, A

ALIGNMENTS

```
RESULT 1
US-09-806-708B-22/c
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.c., L.a., and B.n. PAB1 promoters
US-09-806-708B-22

Query Match          5.4%; Score 51.2; DB 4; Length 1141;
Best Local Similarity 12.2%; Pred. No. 4.1e-05;
Matches 75; Conservative 244; Mismatches 252; Indels 6; Gaps 2;

QY 332 CTGCACTTCACTTATGGAATCACCAGCAAGTACTGCTGATTTAAATATGCT 391
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 729 MTNNNNNNMMWACCTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 670
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 392 TTGAATTCGCGGAAACCATTCGAATGATGAGTTCGATTCGAGGAATATTCAC 451
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 669 RWSTTTRHHHTGATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 610
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 452 TTGTAAGCATGGTGTAAAGACACATGATATGATGATGATGATGATGATGATGAT 511
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 609 DTTCTTVDVWADSVWVWVWVWVWVWVWVWVWVWVWVWVWVWVWVWVWVWVW 550
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 512 GGACAGACTATGACCGCAATCTGCTGTGATTAATCTGATGTCGAGATTCAGAA 571
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 549 WSSWARTTTRNNNNNNNSGBVRMRWAGTMMWNNNNNNNNNNNNNNNNNNNNNNNN 490
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 572 TGCACGTCGAGTAATTCATTCCTGAGCTGCGCAATGTTCCATGAGATGATGAT 631
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 489 NAKSMWGNMNNRRRAKMMWAAANNNDGAMDHWYMGNNNTMWRRAWKKNNNNMWR 430
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 632 TCGACGTCAGTCGATCAAGATATGAAACAGATTTGAAATGATATGATGATGAT 691
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
```


	Matches	75;	Conservative	0;	Mismatches	57;	Indels	0;	Gaps	0;
Qy	819	TCATTATTAAATATCTACATGTAATTTCAATTATCTCTACCTTCGATCTTGCTAT								878
Db	31752	TATTAACATACATATCTTAATATATATTTATTTAAATCTTAACATACATTAATCTAATATATAT								31693
Qy	879	CACATGCTCTTATTTCAAAAATCTCACTTTAAATATTCATATTAATATATGGGTTATTTCA								938
Db	31692	TATTAACCTTAATATATTAACATATCTTAATATATATATATTAATCTTAATATATGTATTAATCT								31633
Qy	939	AATGATCATCT	950							
Db	31632	AATATATATATAT	31621							

```

RESULT 5
US-09-949-016-15597/C
; Sequence 15597, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15597
; LENGTH: 38371
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(38371)
; OTHER INFORMATION: n = A,T,C or G
; US-09-949-016-15597

```

```

; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 94473
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-94473

Query Match      4.2%; Score 40.4; DB 4; Length 601;
Best Local Similarity 53.9%; Pred. No. 0.057;
Matches 83; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY      762 ATTAATCAACAATAATCATCCAGCCTGGAGAGTCGAAAACTGTCGAGCAAGCACTTGCTCT 821
DB      473 ATAAATAAAGTTAAATTTTTCAGCAAGAAATGTTATACCTCGAGGCTGTTTAGATATCT 414

QY      822 TATTTAAATATTTCAACATGTAATTTTCAATTTATCTCTACTTTCTCGATCTTGCTATCAC 881
DB      413 AATTAAATATTTTATTTATATTTATTTTCAGGTTATCTTCCTTCTTATGATTAATTT 354

QY      882 ATGTCCTTATTTCAAAAATCTCACTTTAAATTT 915
DB      353 ATGTCAAATTTTAAAGAAATGATTTTGTATTT 320

```

```

RESULT 7
US-09-949-016-94474/C
; Sequence 94474, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 94474
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-94474

```

```

RESULT 6
US-09-949-016-94473/C
; Sequence 94473, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1.001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
;

```

RESULT 8

```
US-09-949-016-14476/c
; Sequence 14476, Application US/09949016
; Patent No. 6612339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14476
; LENGTH: 20441
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(20441)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14476
```

```
Query Match          4.2%; Score 40.4; DB 4; Length 20441;
Best Local Similarity 53.9%; Pred. No. 0.45;
Matches 83; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
```

```
Oy 762 ATTATCAACAAACCTACACGCTGAGAGTCGAGAACTGCGAGCAAGCACTGGCT 821
Db 14180 ATAAATAAAGTTAATTTTCGAGCAAGAAAGGTTACTGAGCTGTTAGATATCT 14121
```

```
Oy 822 TATTTAATATTTCTACATGTAATTTCAATTTATCTTCTGATCTTGATCAC 881
Db 14120 AATTAAATATTTTATTTATTTATTTTCAGTTATCTTCTTCTGATGATTAAT 14061
```

```
Oy 882 ATGTCTTATTTCAAAAATCTCACTTTAAAT 915
Db 14060 ATGTCAATTTTAAAGAAATGATTTTGTAT 14027
```

```
RESULT 9
US-09-671-317-404/c
; Sequence 404, Application US/09671317
; Patent No. 6528260
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bouguetel, Lydie
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIOMARKERS RELATED TO GENES INVOLVED IN DRUG METABOLISM
; FILE REFERENCE: 62 US3 CIP
; CURRENT APPLICATION NUMBER: US/09/671,317
; CURRENT FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US 09/536,178
; PRIOR FILING DATE: 2000-03-23
; PRIOR APPLICATION NUMBER: PCT/IB00/00403
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: US 60/126,269
; PRIOR FILING DATE: 1999-03-25
; PRIOR APPLICATION NUMBER: US 60/131,961
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 977
; SOFTWARE: Patent.pm
; SEQ ID NO 404
; LENGTH: 1001
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
```

```
; NAME/KEY: allele
; LOCATION: 503
; OTHER INFORMATION: 10-453-330 : polymorphic base C or T
; NAME/KEY: misc.binding
; LOCATION: 483..502
; OTHER INFORMATION: 10-453-330. misl, potential
; NAME/KEY: misc.binding
; LOCATION: 504..523
; OTHER INFORMATION: 10-453-330. mis2, potential complement
; NAME/KEY: primer bind
; LOCATION: 172..169
; OTHER INFORMATION: upstream amplification primer
; NAME/KEY: primer bind
; LOCATION: 578..597
; OTHER INFORMATION: downstream amplification primer, complement
; NAME/KEY: misc.binding
; LOCATION: 491..515
; OTHER INFORMATION: 10-453-330 potential probe
; NAME/KEY: misc.feature
; LOCATION: 19,31,101,212,337,520,661
; OTHER INFORMATION: n=a, g, c or t
US-09-671-317-404
```

```
Query Match          4.2%; Score 40.2; DB 4; Length 1001;
Best Local Similarity 60.6%; Pred. No. 0.089;
Matches 66; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
```

```
Oy 819 TCTTATTTAATATTTCTACATGTAATTTCAATTTATCTTCTGATCTTGCTAT 878
Db 802 TTTAATTAACATTTTATTTATTTATTTCAAGATTTTCTCTTCTACTATGATTC 743
```

```
Oy 879 CACATGCTCTTATTTCAAAAATCTCACTTTAAATTCATTAATTAAT 927
Db 742 ATTAGTCAATTTTAAAGAAATGATTTTATTTATTTATCTATATCAT 694
```

```
RESULT 10
US-09-949-016-16778/c
; Sequence 16778, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16778
; LENGTH: 15236
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16778
```

```
Query Match          4.2%; Score 40; DB 4; Length 15236;
Best Local Similarity 53.9%; Pred. No. 0.5;
Matches 82; Conservative 0; Mismatches 70; Indels 0; Gaps 0;
```

```
Oy 795 GAAAACTGTCGAGCAAGCACTGCTTATTTAAATTTCTAATCAATTTCAATTTA 854
Db 4646 GAAAAACATCTAGAGCAGATTTAATTAATTAATGTTAATTTCTTTCTATC 4587
```

```
Oy 855 TCTCTACTTCTGATCTGCTATCAATGCTCTTATTTCAAAATCTCACTTTAAAT 914
Db 4586 TGTATCTTTATTTTATTAAGAAATTAACATGCTTTTATTAAGAAATTAATTTCTGAAT 4527
```

Qy 915 TCATATAATATGGTTTATTCAAATACATC 946
Db 4526 TAAAGAAAAAAGTGTATATTAAAAAGCAGC 4495

RESULT 11

US-09-790-988-1
Sequence 1, Application US/09790988

Patent No. 6632335
GENERAL INFORMATION:
APPLICANT: SHIGENOBU, SHUJI
APPLICANT: MATSUNABE, HIDEKI
APPLICANT: HATTORI, MASAHIRA
APPLICANT: SAKAKI, YOSHIYUKI
TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
FILE REFERENCE: 08136/0159
CURRENT FILING DATE: 2001-02-23
PRIORITY FILING DATE: 2000-04-07
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 640681
TYPE: DNA
ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match 4.2%; Score 40; DB 4; Length 640681;
Best Local Similarity 52.4%; Pred. No. 4.5;
Matches 88; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy 780 CCAGCTGAGAGTCGAAACTGTCGAGACGACCTTGCTTATTTAAATTTCTTACA 839
Db 143181 CCAAGCAGAAATATGCAAAACATCTTATTAATTTAAATTTTATGACTATTTAAAAA 143240
Qy 840 TGTATTTCAATTTATCTTACTTCTGATCTTGCTATCAAGCTCTTATTTCAAAA 899
Db 143241 TATTTTTCATTTTAAAGAAATATTTTAAATTTTAAATTTTAAAGTATGCTTTTAAAA 143300
Qy 900 ATCTCACTTTAAATTCATATTAATATGAGTTTATTTCAATTCATCA 947
Db 143301 TTTTTCCTTTAAAAAAGATTAAGAAATCTTATTTAAAAAATACA 143348

RESULT 12

US-09-949-016-11820
Sequence 11820, Application US/09949016

Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: C1001307
CURRENT FILING DATE: 2000-04-14
PRIORITY FILING DATE: 2000-10-20
PRIORITY FILING DATE: 2000-10-20
PRIORITY FILING DATE: 2000-10-03
PRIORITY FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11820
LENGTH: 192700
TYPE: DNA
ORGANISM: Human
US-09-949-016-11820

Query Match 4.2%; Score 39.6; DB 4; Length 192700;
Best Local Similarity 54.3%; Pred. No. 2.9;
Matches 101; Conservative 0; Mismatches 84; Indels 1; Gaps 1;

Qy 759 AGCATTAATCAACAATCTACAGCTGGAGAGTCGAAAACTGTCGAGCAAGCACTTG 818
Db 41097 AGTAATATTTATTAATTAATTCAGTACGACTAAAGAAAAATACCAAACTAAAGCAATTTTG 41156
Qy 819 TCTTATTTAAATATCTTAACATGTAATTTCAATTTATCTTACTTCTGATCTTAT 878
Db 41157 AATTAGCATTTGAATCTAA-AAGAAAGTAGATTTTTTCTGAAATCTGTTCTAGAGTGT 41215
Qy 879 CACATGCTCTTATTTCAAAAACTCACTTTAAATTCATATTAATATGAGTTTATCA 938
Db 41216 GGTATATGTATTTTTCAGAAAACTTAAACAAATTTGCGCAAAATGAAGAAATATTTA 41275
Qy 939 AATACA 944
Db 41276 AAAACA 41281

RESULT 13

US-09-949-016-17182
Sequence 17182, Application US/09949016

Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: C1001307
CURRENT FILING DATE: 2000-04-14
PRIORITY FILING DATE: 2000-10-20
PRIORITY FILING DATE: 2000-10-20
PRIORITY FILING DATE: 2000-10-20
PRIORITY FILING DATE: 2000-10-03
PRIORITY FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 17182
LENGTH: 192704
TYPE: DNA
ORGANISM: Human
US-09-949-016-17182

Query Match 4.2%; Score 39.6; DB 4; Length 192704;
Best Local Similarity 54.3%; Pred. No. 2.9;
Matches 101; Conservative 0; Mismatches 84; Indels 1; Gaps 1;

Qy 759 AGCATTAATCAACAATCTACAGCTGGAGAGTCGAAAACTGTCGAGCAAGCACTTG 818
Db 41097 AGTAATATTTATTAATTAATTCAGTACGACTAAAGAAAAATACCAAACTAAAGCAATTTTG 41156
Qy 819 TCTTATTTAAATATCTTAACATGTAATTTCAATTTATCTTACTTCTGATCTTAT 878
Db 41157 AATTAGCATTTGAATCTAA-AAGAAAGTAGATTTTTTCTGAAATCTGTTCTAGAGTGT 41215
Qy 879 CACATGCTCTTATTTCAAAAACTCACTTTAAATTCATATTAATATGAGTTTATCA 938
Db 41216 GGTATATGTATTTTTCAGAAAACTTAAACAAATTTGCGCAAAATGAAGAAATATTTA 41275
Qy 939 AATACA 944
Db 41276 AAAACA 41281

RESULT 14

US-09-949-016-17613
Sequence 17613, Application US/09949016

Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
FILE REFERENCE: C1001307

;/ CURRENT APPLICATION NUMBER: US/09/949,016
;/ CURRENT FILING DATE: 2000-04-14
;/ PRIOR APPLICATION NUMBER: 60/241,755
;/ PRIOR FILING DATE: 2000-10-20
;/ PRIOR APPLICATION NUMBER: 60/237,768
;/ PRIOR FILING DATE: 2000-10-03
;/ PRIOR APPLICATION NUMBER: 60/231,498
;/ PRIOR FILING DATE: 2000-09-08
;/ NUMBER OF SEQ ID NOS: 207012
;/ SOFTWARE: FastSeq for Windows Version 4.0
;/ SEQ ID NO: 17613
;/ LENGTH: 33044
;/ TYPE: DNA
;/ ORGANISM: Human
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (1)...(33044)
;/ OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17613

Query Match 4.0%; Score 38.6; DB 4; Length 33044;
Best Local Similarity 58.1%; Pred. No. 2.1;
Matches 68; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 804 CGAGCAAGCCTGCTCTATTATTAATTTCTACATGTAATTCATTATCTCTACT 863
DB 6484 CTAGCTAACATTTGGCTTTAATAATGTTTAAATTTGGTGTTGTTTTCTTACA 6543
QY 864 TTCTGATCTTGCTATCACATGCTCTTATTGAAAAATCTCACTTTAAATTCATAT 920
DB 6544 TTCTTTAGACATGCTCTGTGCTGCGCAAAAGTAAGATCATATGTTACAT 6600

RESULT 15

US-09-949-016-12455
;/ Sequence 12455, Application US/09949016
;/ Patent No. 6812339
;/ GENERAL INFORMATION:
;/ APPLICANT: VENTER, J. Craig et al.
;/ TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;/ FILE REFERENCE: C1001307
;/ CURRENT APPLICATION NUMBER: US/09/949,016
;/ CURRENT FILING DATE: 2000-04-14
;/ PRIOR APPLICATION NUMBER: 60/241,755
;/ PRIOR FILING DATE: 2000-10-20
;/ PRIOR APPLICATION NUMBER: 60/237,768
;/ PRIOR FILING DATE: 2000-10-03
;/ PRIOR APPLICATION NUMBER: 60/231,498
;/ PRIOR FILING DATE: 2000-09-08
;/ NUMBER OF SEQ ID NOS: 207012
;/ SOFTWARE: FastSeq for Windows Version 4.0
;/ SEQ ID NO: 12455
;/ LENGTH: 69813
;/ TYPE: DNA
;/ ORGANISM: Human
US-09-949-016-12455

Query Match 4.0%; Score 38.6; DB 4; Length 69813;
Best Local Similarity 45.5%; Pred. No. 3.3;
Matches 137; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

QY 650 CAAGTATGAACACAGATTGAAAAATGAATCATTCGTCATTGGAAACAAGTCAAAAAG 709
DB 64176 CAGAGATATTAATTAAATAAAGAAACACTATATAAAATATGAAAAATTTATGAG 64235
QY 710 GTTGTGCTCATCAAGCTGCTCGACTGCTCCCATGCAAGACATAGCATTAATCA 769
DB 64236 CTATGTAATCATATCAACATACCTCAATATAGAAAACAAAATTTATAAATTTATAG 64295
QY 770 ACAAATCTACAGCCTGGAAGAGTGAAAACTGTCGAGCAAGCACTGGCTTATTTAA 829
DB 64296 AATTAATCTACAGATGAAAAATTTCAATATATCTTGTCCCAAAATGAATTTAAATACA 64355

QY 830 TATTCTAACAGTAATTTCAATTTATCTCTACTTTCGATCTTGCTATCAGATGCTCT 889
DB 64356 GACAAAATGAGTGATATATAGATATTTCACTCAAAATTTGACAGACTAATTTGAT 64415
QY 890 TATTCAAAAATCTCACTTTAAATTCATATATAATGAGTTTATTCAAATACATCATC 949
DB 64416 ATACAGATTAATGTATATTTCTGTAAACATATATATCTGTATGAATCTATTAATACATC 64475
QY 950 T 950
DB 64476 T 64476

Search completed: April 9, 2005, 06:21:45
Job time : 208 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 9, 2005, 01:17:04 ; Search time 615 Seconds

(without alignments)
9182.818 Million cell updates/sec

Title: US-09-087-136-2

Perfect score: 954

Sequence: 1 atgtcagaatagatccact.....ttcaatcacatcttgac 954

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: _geneseqn1980s:*
2: _geneseqn1990s:*
3: _geneseqn2000s:*
4: _geneseqn2001as:*
5: _geneseqn2001bs:*
6: _geneseqn2002as:*
7: _geneseqn2002bs:*
8: _geneseqn2003as:*
9: _geneseqn2003bs:*
10: _geneseqn2003cs:*
11: _geneseqn2003ds:*
12: _geneseqn2004as:*
13: _geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	952.4	99.8	954	2	AAV72859
2	47.6	5.0	2000	8	ADA71938
3	45.6	4.8	427	12	ADM67000
4	44.6	4.7	6291	6	ABL34038
5	44.6	4.7	6291	6	ABK31482
6	44.4	4.7	6292	4	AA646795
7	44.4	4.5	8346	6	ABK28327
8	42.6	4.5	50000	6	ABL55643
9	42.4	4.4	6874	4	AA645440
10	42.4	4.4	6874	4	ABL70381
11	42.4	4.4	6874	6	ABK28289
12	42.4	4.4	6874	6	AA61333
13	42.4	4.4	6874	6	ABN80230
14	41.4	4.3	129017	12	ADP84158
15	41.2	4.3	15320	4	ABL07834
16	41.1	4.3	5511	6	ABL33871
17	40.4	4.2	6380	6	ABN80109
18	40.2	4.2	1001	3	AAH51513
19	40.2	4.2	7560	6	ABL33223
20	40.2	4.2	15767	6	ABL33206

C 21	40.2	4.2	15767	6	ABL34552	AB134552 Human met
C 22	40.2	4.2	15767	7	ADS98813	Ad998813 Bimulphit
C 23	40.0	4.2	11780	4	AAK71099	AAK71100 Human imm
C 24	40.0	4.2	16299	4	AAK71100	AAK71100 Human imm
C 25	40.0	4.2	17131	6	ABL33053	AB133053 Human imm
C 26	40.0	4.2	31236	9	ADA02900	Ada02900 Human PTP
C 27	40.0	4.2	31236	10	AD872638	Ad872638 Human PTP
C 28	40.0	4.2	31236	10	ADC85379	Adc85379 Mouse Ptp
C 29	40.0	4.2	31236	12	ADM74495	Adm74495 Human car
C 30	40.0	4.2	110000	6	ABA92787_1	Continuation (2 of
C 31	39.8	4.2	5641	6	ABL33396	AB133396 Human imm
C 32	39.8	4.2	7892	6	ABK40056	ABK40056 Human che
C 33	39.6	4.2	208700	13	ABD32688	ABd32688 Human can
C 34	39.4	4.1	273	4	AA336231	AA336231 Human car
C 35	39.4	4.1	273	4	AA336230	AA336230 Human car
C 36	39.4	4.1	273	10	ADE46924	Ad46924 Human car
C 37	39.4	4.1	273	10	ADE46925	Ad46925 Human car
C 38	39.4	4.1	273	13	ADJ08342	Adj08342 Human car
C 39	39.4	4.1	273	13	ADJ08343	Adj08343 Human car
C 40	39.4	4.1	282	5	ABV47775	ABv47775 Human pro
C 41	39.2	4.1	12781	4	AA645386	AA645386 Tumour su
C 42	39.2	4.1	12781	6	ABL92229	AB192229 Chemical
C 43	39.2	4.1	12781	10	ADB54128	ABd54128 Pretreat
C 44	39.0	4.1	619	5	ABV04737	ABv04737 Human pro
C 45	38.8	4.1	12781	10	ADB54256	ABd54256 Pretreat

ALIGNMENTS

RESULT 1	AAV72859	standard; cDNA; 954 BP.
ID	AAV72859	
XX	AAV72859;	
AC	29-MAR-1999	(first entry)
DT	Caenorhabditis elegans bynmuv gene lin-37 cDNA.	
XX	Caenorhabditis elegans bynmuv gene lin-37 cDNA.	
XX	lin-37; synthetic multivulvar; bynmuv; signal transduction; animal model;	
KW	tumour suppressor; retinoblastoma; cancer; cancer; cell proliferation;	
KW	gene therapy; ss.	
OS	Caenorhabditis elegans.	
XX		
FX	Key	Location/Qualifiers
FT	CDS	1..828
FT		/*tag= a
XX	W09854299-A1.	
XX	03-DEC-1998.	
XX	28-MAY-1998;	98MO-US011043.
XX	28-MAY-1997;	97UG-0047996P.
PA	(MASI) MASSACHUSETTS INST TECHNOLOGY.	
XX	Hovrlitz HR, Ceol C, Lu X;	
XX	WPI, 1999-045362/04.	
DR	P-PSDB; AAW83386.	
XX	Novel lin-37, -35, -55, -52, -53 and -54 gene from C. elegans - useful	
PT	for treating diseases associated with altered levels of cell	
PT	proliferation, e.g. carcinomas.	
XX	Claim 10; Fig 3; 70pp; English.	
PS	This cDNA sequence codes for lin-37 (see AAW83386), a novel 32 kDa	
CC	hydrophilic protein of Caenorhabditis elegans. lin-37 cDNA was isolated	

CC From the Okkema embryonic cDNA library. The lin-37 transcript is about 1
CC kb in size and is present both in embryonic and mixed-stage RNAs. The lin
CC -35 gene is a novel synthetic multiviral (SynMuv) gene involved in cell
CC fate and cell proliferation, and is part of a pathway that may be used as
CC a genetic and biochemical model system for tumour suppression and cancer
CC in mammals. SynMuv pathway genes and proteins may be used to identify
CC genes which are part of the mammalian pathway and to identify genes,
CC proteins and therapeutic compounds which modulate this pathway. Pure
CC nucleic acids (see AAV72859-65) encoding C. elegans SynMuv polypeptides
CC selected from LIN-37, -35, -55, -52, -53, -54 and 52F-1 (see AAM83386-
CC 92), are new. Also claimed are: (1) nucleic acids (see AAV72866-67)
CC encoding the mouse and human homologues of C. elegans LIN-54; (2) vectors
CC containing the nucleic acids; (3) transgenic cells; (4) a pure mammalian
CC SynMuv polypeptide; and (5) an antibody which binds to a SynMuv family
CC protein. The SynMuv nucleic acids and polypeptides can be used to
CC diagnose and treat, especially by gene therapy, conditions involving
CC altered levels of cell proliferation, e.g. SynMuv-associated carcinomas
XX

Query Match 99.8%; Score 952.4; DB 2; Length 954;
Best Local Similarity 99.9%; Pred. No. 4,4e-247;

Matches 953; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGTGGAATATGATCCAGTTCGCGAGTCTTGTCTTCCAGAGAGGAGATGAAATGCT 60
DB 1 ATGTGGAATATGATCCAGTTCGCGAGTCTTGTCTTCCAGAGAGGAGATGAAATGCT 60
QY 61 CGTCAAAATGATCCATTGATTAAGCGAGGTCACCTTCATTGGAATTCGCAAGCAAAA 120
DB 61 CGTCAAAATGATCCATTGATTAAGCGAGGTCACCTTCATTGGAATTCGCAAGCAAAA 120
QY 121 CTCACATCCCGTGTATCCATATATCCGACAGTTCGCGAGTCCCGGATTAATTCGCGC 180
DB 121 CTCACATCCCGTGTATCCATATATCCGACAGTTCGCGAGTCCCGGATTAATTCGCGC 180
QY 121 CTCACATCCCGTGTATCCATATATCCGACAGTTCGCGAGTCCCGGATTAATTCGCGC 180
DB 121 CTCACATCCCGTGTATCCATATATCCGACAGTTCGCGAGTCCCGGATTAATTCGCGC 180
QY 181 AGAAAAAGTCTGGGAATCTGCTGACCAACATATAACATCCATCCGAAATATATGGA 240
DB 181 AGAAAAAGTCTGGGAATCTGCTGACCAACATATAACATCCATCCGAAATATATGGA 240
QY 181 AGAAAAAGTCTGGGAATCTGCTGACCAACATATAACATCCATCCGAAATATATGGA 240
DB 181 AGAAAAAGTCTGGGAATCTGCTGACCAACATATAACATCCATCCGAAATATATGGA 240
QY 241 GTACTCCCAAGAGATTATATCTGCTGATGAAGAGCCCGGCGCAAGAGCTGCACCA 300
DB 241 GTACTCCCAAGAGATTATATCTGCTGATGAAGAGCCCGGCGCAAGAGCTGCACCA 300
QY 301 GGTGCGCCCTCGTAAGTCCGCGACAGATCTTCACTTCACTTAATGATTCACCAAGC 360
DB 301 GGTGCGCCCTCGTAAGTCCGCGACAGATCTTCACTTCACTTAATGATTCACCAAGC 360
QY 301 GGTGCGCCCTCGTAAGTCCGCGACAGATCTTCACTTCACTTAATGATTCACCAAGC 360
DB 301 GGTGCGCCCTCGTAAGTCCGCGACAGATCTTCACTTCACTTAATGATTCACCAAGC 360
QY 361 AAGACTATGACTCGTATCTTAATAATTAATGTTGAATGGCGTGAAGAACATTCGAAATG 420
DB 361 AAGACTATGACTCGTATCTTAATAATTAATGTTGAATGGCGTGAAGAACATTCGAAATG 420
QY 421 ATAGTGGACGTTTGAAGAGAAATATTCACCTTGTATAGCATGGGTTAAAGACACATG 480
DB 421 ATAGTGGACGTTTGAAGAGAAATATTCACCTTGTATAGCATGGGTTAAAGACACATG 480
QY 481 AATAATGAATATGAACCAATTAAGCTCAAGAGACACTATGACCGAATCTGGCTGT 540
DB 481 AATAATGAATATGAACCAATTAAGCTCAAGAGACACTATGACCGAATCTGGCTGT 540
QY 481 AATAATGAATATGAACCAATTAAGCTCAAGAGACACTATGACCGAATCTGGCTGT 540
DB 481 AATAATGAATATGAACCAATTAAGCTCAAGAGACACTATGACCGAATCTGGCTGT 540
QY 541 GATTATCTTGGATGCGGAGATTCATGCAATGCGACGTCGAGTAATCAATTCCTGAG 600
DB 541 GATTATCTTGGATGCGGAGATTCATGCAATGCGACGTCGAGTAATCAATTCCTGAG 600
QY 601 CTGCAATTTGTTCCATCTAGATGATGAATTCGACGTTAAGTGCATCAAGATATGAA 660
DB 601 CTGCAATTTGTTCCATCTAGATGATGAATTCGACGTTAAGTGCATCAAGATATGAA 660
QY 661 ACAGATTTGAATAATGATATGATTCGATTCGATGGAAGCAAGCAAAAAGTTGGTGGCT 720
DB 661 ACAGATTTGAATAATGATATGATTCGATTCGATGGAAGCAAGCAAAAAGTTGGTGGCT 720
QY 721 CATCAAGCTGCTCGAGTGTCTCCCATGCAAGAAACATGATTAATCAAAAATCTAC 780

DB 721 CATCAAGCTGCTCGAGTGTCTCCCATGCAAGAAACATGATTAATCAAAAATCTAC 780
QY 781 CAGCTGAGAGATGAGAAACCTGTCGACCAAGCACTGTATTTAATTTAATCTACAT 840
DB 781 CAGCTGAGAGATGAGAAACCTGTCGACCAAGCACTGTATTTAATTTAATCTACAT 840
QY 841 GTAATTTCAATTTATCTTACTTCTTGTGATCTGCTATTCACAGTCTCTTATTTCAAAA 900
DB 841 GTAATTTCAATTTATCTTACTTCTTGTGATCTGCTATTCACAGTCTCTTATTTCAAAA 900
QY 901 TCTCAGTTTAAATTCATATTAATTAATGAGTTTATTCGAATTCATCTGAC 954
DB 901 TCTCAGTTTAAATTCATATTAATTAATGAGTTTATTCGAATTCATCTGAC 954

RESULT 2

ADA71938 standard; DNA; 2000 BP.

ADA71938;

20-NOV-2003 (first entry)

Rice gene, SEQ ID 5263.

plant; bacterial infection; fungal infection; viral infection; rice;

gene; ds.

Oryza sativa.

WO2003000898-A1.

03-JAN-2003.

22-JUN-2001; 2001WO-IB001105.

22-JUN-2001; 2001WO-IB001105.

(SYGN) SYNGENTA PARTICIPATIONS AG.

Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;

Katagiri F, Quan S, Tao Y, Whitman S, Xie Z, Zhu T, Zou G;

WPI; 2003-175290/17.

Identifying at least one gene involved in plant resistance or response to

pathogenic infection for conferring resistance or tolerance to a plant to

bacterial, fungal or viral infection by determining or detecting plant

gene expression.

Claim 27; SEQ ID NO 5263; 899bp; English.

The present invention relates to a method (M1) for identifying genes

involved in plant resistance or response to pathogenic infection. M1

comprises identifying a gene whose expression is significantly altered in

the incompatible interaction of plant gene expression relative to

expression of the gene in an uninfected plant, in a mutant plant that

does not express a gene associated with response to pathogenic infection,

or in a corresponding incompatible or compatible interaction. (M1) is

useful for conferring resistance to resistance or tolerance to a plant to

bacterial, fungal or viral infection. The present sequence was used to

illustrate the invention.

Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match 5.0%; Score 47.6; DB 8; Length 2000;

Best Local Similarity 8.7%; Pred. No. 0.034;

Matches 26; Conservative 154; Mismatches 118; Indels 0; Gaps 0;

QY 653 GATATGAAACAGATTGAAATAATGATATGATTCGATTCGATTCGAAACAGTCAAAAAGTT 712
DB 501 RRRRWRMR 560

[illegible]

hypotensive activities. This polynucleotide sequence is a murine adipocyte specific DNA sequence of the invention.

Sequence 427 BP; 153 A; 60 C; 85 G; 129 T; 0 U; 0 Other;

Query Match 4.8%; Score 45.6; DB 12; Length 427;
Best Local Similarity 59.1%; Pred. No. 0.062; Mismatches 54; Indels 0; Gaps 0
Matches 78; Conservative 0;

Oy 820 CTATTAAATTTCTAACATGTGTAATTTCAATTTAATTCCTTACTTGATCTTGCTATC 879
Db 164 CTACGTAGAGATGTGTAATTTCAATTCCTTTTAAATTTCTTTCTTTCTTTCTTTCTTC 105

Oy 880 ACATGCTCTTATTTCAAATAATCTCACCTTAAAATTCATATAAATAGCGTTTATTCAA 939
Db 104 TCCTTTTTTTTTTTCAAATAATCAAACGTAAAAAACTTTCAACGTCACGATATTTAA 45

Oy 940 ATACATCATCTT 951
Db 44 TTAAAGTACTT 33

RESULT 4
ABL34038/C
ID ABL34038 standard; DNA; 6291 BP.
XX
XX
AC ABL34038;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 2011.
XX
XX
Human; immune system disease; cytosine methylation; antiaesthetic;
antiartherosclerotic; antihaemic; cyostatic; nocropic;
neuroprotective; anti-HIV; anticoagulant; ophthalmological;
antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
ds.
KW
KM
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-BP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
XX
PS methylation.
XX
XX
Claim 1; SEQ ID NO 2011; 32pp + Sequence Listing; German.

The present invention provides a number of human immune system associated genes which are modified by the methylation of cytosines. The sequences can be used in the diagnosis and treatment of immune system disorders, including eye diseases such as retinopathy, neovascular glaucoma and macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis, rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel diseases. The present sequence is a gene of the invention

with signal transduction, or their complementary sequences. Note: The

CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The

CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 6292 BP; 1488 A; 213 C; 1576 G; 3015 T; 0 U; 0 Other;

Query Match 4.7%; Score 44.6; DB 4; Length 6292;
Best Local Similarity 58.8%; Pred. No. 0.35;
Matches 77; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 812 CACTTGCTTATTTAAATATCTTAACATGTAATTTCAATTTATCTTACTTCTGATC 871

DB 3386 CATTATATCTTACATTAATTAATCTTCAATCATCTCATATCTTCTCTTAT 3327

QY 872 TTGCTATCATGTCCTCTTATTTCAAAAATCTCACTTTAAATTCATTAATAATGCGT 931

DB 3326 ATCCCATAAATATCTATTTCTCAAAAAAATTTATTTAAACCTCATTAATAACAC 3267

QY 932 TTATTCAAATA 942

DB 3266 TTTTAAACA 3256

RESULT 7

ABK28327 ID ABK28327 standard; DNA; 8346 BP.

XX AC ABK28327;

DT 23-APR-2002 (first entry)

DE DNA transcription associated genomic DNA #101.

XX DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KM PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KM single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KM viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KM immunological disorder; Werner syndrome; developmental disorder;
KM psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
KM neurodegenerative disorder; Mardenburg syndrome; Niemann-Pick disease;
KM myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KM angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
KM polyglutamine disorder; solid tumour.

XX Unidentified.

OS WO200192565-A2.

PN 06-DEC-2001.

PD 06-APR-2001; 2001WO-EP003973.

XX 06-APR-2000; 2000DE-01019058.

PR 07-APR-2000; 2000DE-01019173.

PR 30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-090046/12.

XX New nucleic acids or oligomers, useful for diagnosing or treating

PT diseases associated with DNA transcription, e.g. immunological disorders,

PT Werner syndrome, psoriasis, myocardial infarction, solid tumours or
PT cancer.

PS Claim 1; SEQ ID NO 201; 32pp; English.

XX The invention relates to a nucleic acid, which comprises a segment of the
CC chemically pretreated DNA of genes associated with DNA transcription from
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC to the chemically pretreated DNA of genes associated with DNA
CC transcription. The set of oligomer probes are useful for detecting the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC diagnosing or treating diseases associated with DNA transcription
CC (particularly with the methylation status), e.g. adenosine deaminase
CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC neurological disorders, neurodegenerative disorders, Mardenburg
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC associated genomic DNA molecules of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification but
CC was obtained in electronic format directly from the European Patent
CC Office

XX Sequence 8346 BP; 2154 A; 239 C; 1768 G; 4185 T; 0 U; 0 Other;

Query Match 4.7%; Score 44.4; DB 6; Length 8346;
Best Local Similarity 59.5%; Pred. No. 0.45; Indels 0; Gaps 0;
Matches 75; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 817 GGTCTTATTTAAATATCTTAACATGTAATTTCAATTTATCTTACTTGTGATCTTGGT 876

DB 3653 GGTATATTTTAAAGTTATTTTGTGTTTGTGTTTGTGTTTATGATTTGGGA 3712

QY 877 ATCAGATGCTCTTATTTCAAAAATCTCACTTTAAATTCATTAATAATGCGTTAT 936

DB 3713 TTATGATTTTATTTTAAATTTTATTTTATTTAGTTAGATGATGATTTTGTA 3772

QY 932 CAAATA 942

DB 3773 TAAATA 3778

RESULT 8

ABL55643 ID ABL55643 standard; DNA; 5000 BP.

XX ABL55643;

DT 01-JUL-2002 (first entry)

DE AMLPV genome fragment#1.

XX AMLPV; gene therapy; viral vector; chromosome mapping; gene mapping;

XX genetic deficiency disorder; ds.

OS Ammacta moorei entomopoxvirus.

PN WO200212526-A2.

PD 14-FEB-2002.

XX 10-AUG-2001; 2001WO-US025287.

PR 10-AUG-2000; 2000US-0224479P.

PR 14-SEP-2000; 2000US-0062254.

XX (UVFL) UNIV FLORIDA.

PI Moyer RM, Li Y, Bawden AL;
XX WPI; 2002-227161/28.
XX
XX Novel recombinant entomopox virus vector useful for delivering
PT polynucleotide encoding protein to vertebrate cell, comprises
PT polynucleotide encoding protein operably linked with heterologous
PT promoter sequence.
XX
XX Disclosure; Page 125-150; 326pp; English.
XX
XX The invention relates to a recombinant entomopox virus (BPV) vector,
CC comprising a polynucleotide encoding a protein operably linked with a
CC heterologous promoter sequence. The invention also concerns methods for
CC providing gene therapy for genetic deficiency disorders. Vectors of the
CC invention are useful for delivering a polynucleotide encoding a protein
CC to a vertebrate cell preferably a mammalian cell, such as a human cell.
CC The vector is introduced into the vertebrate cell by infection in a viral
CC particle, or by transfection, transduction, or infection either in vitro
CC or in vivo. The vector is useful for the delivery and expression of
CC biologically useful proteins in gene therapy protocols, and for
CC delivering large DNA segments for engineering of vertebrate cells.
CC Polynucleotides of the invention have applications in techniques such as
CC their use as insertion sites for foreign genes of interest, hybridisation
CC probes, for chromosome and gene mapping, in PCR technologies, and in the
CC production of sense or antisense nucleic acids. Vectors of the invention
CC provide for stable integration and expression of heterologous DNA in host
CC cells, and are adapted for accepting large heterologous polynucleotide
CC inserts which can be delivered in an infected or transformed cell and
CC expressed in a stable fraction. The current sequence represents a
CC fragment of the genome of the genus B entomopoxvirus from *Amsacta moorei*
CC (AMBPV)
XX
XX Sequence 50000 BP; 19857 A; 4373 C; 4222 G; 21548 T; 0 U; 0 Other;
SQ

Query Match 4.5%; Score 42.6; DB 6; Length 50000;
Best Local Similarity 61.1%; Pred. No. 2.9;
Matches 69; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 815 TTGGCTTATTAAATATTCACATGTAATTCATTATCTTACTTCTGATCTTG 874
DB 35961 TTGCGCGCTCTAAATTAATGTTGATTTAAATTTATATGCTCATTTATTTATTT 36020
QY 875 CTATCAATGCTCTTATTTCAAAATCATCTTAATAATCATTAATAATAT 927
DB 36021 TGATATCAATTCATTTCCAAAATTTTCATTATTAATAATATTAATAT 36073

RESULT 9
AAS45440/C
ID AAS45440 standard; DNA; 6874 BP.
XX
XX AAS45440;
AC
XX
XX 18-DEC-2001 (first entry)
DT
XX
XX Chemically pretreated genomic DNA associated with cell cycle #73.
DE
XX Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
XX human immunodeficiency virus; neurodegenerative disorder; solid tumour;
XX graft-versus-host disease; glomerular disease; lewy body disease; cancer;
XX arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
XX immunosuppressive; antitumour; cytoskeletal; antiarteriosclerotic; ds;
XX PCR primer.
XX
XX Homo sapiens.
OS
XX
XX WO200168911-A2.
PN
XX
XX 20-SEP-2001.
PD
XX
XX 15-MAR-2001; 2001WO-EP002945.
PF
XX
XX

PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIG-) EPIGENOMICS AG.
PA
XX
XX Olek A, Piepenbrock C, Berlin K;
PI
XX WPI; 2001-602751/68.
DR
XX
XX Designing primers and probes for analyzing diseases associated with
PT cytosine methylation state e.g. arthritis, cancer, aging,
PT arteriosclerosis comprising fragments of chemically modified genes
PT associated with cell cycle.
XX
XX Claim 1; SEQ ID NO 145; 28pp; English.
XX
XX Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
CC molecules associated with the cell cycle and specific PCR primers of the
CC invention. The sequences are useful for detecting the methylation state
CC of all CpG dinucleotides in a sequence and therefore for analyzing
CC associated diseases. By analysing cytosine methylations in the pretreated
CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC of existing diseases or the predisposition to specific diseases can be
CC ascertained. The parameters may be compared to another set of genetic
CC and/or epigenetic parameters, the differences serving as basis for
CC diagnosis and/or prognosis events which are disadvantageous to patients.
CC The sequences of the invention are useful for the diagnosis and therapy
CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC aging, glomerular disease, lewy body disease, arthritis,
CC arteriosclerosis, solid tumours and cancers
XX
XX Sequence 6874 BP; 1589 A; 259 C; 1935 G; 3091 T; 0 U; 0 Other;
SQ

Query Match 4.4%; Score 42.4; DB 4; Length 6874;
Best Local Similarity 50.5%; Pred. No. 1.4;
Matches 103; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 747 TGCAAGACATGACATTAATCAACAATCTACACCTGGAAGCGAAACTGCGA 806
DB 2842 TACATTTATATTAATTAACCTTACAAATTTACATTTATTAATCAAAAAATTAAT 2783
QY 807 GCAAGCACTGGCTCTTATTTAAATATTCATGTAATTCATTTATCTTACTTTC 866
DB 2782 ATTATCAATTTCTATATATTCATATATATATATTAATTAACAATTTTACATCA 2723
QY 867 TGATCTGCTATGACATGCTCTTATTTCAAAAATCTCACTTTAAATCTATTAATA 926
DB 2722 CTACTATTAATTCACAAATCCAAAACCTTAAAAAATTAATAATTAATTAACAATA 2663
QY 927 TGGGTTATTCAAATTAACATCATCT 950
DB 2662 TCAATAAAATTAATTAATTAATTTT 2639

RESULT 10
ABL70381/C
ID ABL70381 standard; DNA; 6874 BP.
XX
XX ABL70381;
AC
XX
XX 01-JUN-2002 (first entry)
DT
XX
XX Chemically treated cell signalling DNA sequence#136.
DE
XX Cell signalling; cytosine methylation; cell signalling disease; cancer;
XX tumour; cytoskeletal; ds.
XX
XX Unidentified.
OS
XX
XX WO200202807-A2.
PN
XX
XX

XX 10-JAN-2002.
XX 29-JUN-2001; 2001WO-EP007471.
XX 30-JUN-2000; 2000DE-01032529.
XX 01-SEP-2000; 2000DE-01043826.
XX (EPIC-) EPIDENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-154758/20.
XX Nucleic acid, useful for diagnosis and therapy of diseases associated
PT with cell signalling e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with cell signalling.
XX
XX Claim 1; SEQ ID NO 271; 24bp + Sequence Listing; English.
XX
XX The invention relates to a nucleic acid comprising a sequence of at least
CC 18 bases of a segment of chemically pretreated DNA of genes associated
CC with cell signalling. The activity of the modified sequences of the
CC invention may be described as cytostatic. The object of the invention is
CC to provide the chemically modified DNA of genes associated with cell
CC signalling, as well as oligonucleotides and/or RNA-oligonucleotides for
CC detecting cytosine methylations, as well as a method which is
CC particularly suitable for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with cell signalling. The
CC chemically modified DNA provided by the invention is useful for diagnosis
CC and therapy of diseases such as solid tumours and cancer. The sequences
CC given in records AB170111-AB170626 represent chemically pre-treated
CC genomic DNA's of genes associated with cell signalling. Note: The
CC sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office
XX
SQ Sequence 6874 BP; 1589 A; 259 C; 1935 G; 3091 T; 0 U; 0 Other;
Query Match 4.4%; Score 42.4; DB 6; Length 6874;
Best Local Similarity 50.5%; Pred. No. 1.4;
Matches 103; Conservative 0; Mismatches 101; Indels 0; Gaps 0;
QY 747 TCGAAGACATGATTAATCAACAAATCTACAGCTGGAAGTGGAAATCTGCA 806
DB 2842 TACATTATATTAATTAACCCCTACAAATTCATTTTAAATCAAAAAATTAAT 2783
QY 807 GCAAGCACTGCTTATTAATTAATCTAATGTAATTCATTTTCTTACTTTC 866
DB 2782 ATTATCAATTTCTCATATTAATTAATTAATTAATTAATTAATTAATTAAT 2723
QY 867 TGAATCTGATCAATGCTCTTATTTCAAAATCTCACTTTAAATCAATTAATTA 926
DB 2722 CTTCATTAATTCACAAATCAAAATCTTAATAAATTAATTAATTAATTAATTA 2663
QY 927 TGGGTTTATTAATTAATTAATCTATCT 950
DB 2662 TCAATTAATTAATTAATTAATTTT 2639
RESULT 11
ABK28289/c
ID ABK28289 standard; DNA; 6874 BP.
XX
XX ABK28289;
XX
XX 23-APR-2002 (first entry)
XX
XX DNA transcription associated genomic DNA #82.
XX
XX DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KM PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KM single nucleotide polymorphism; adenosine deaminase deficiency; cancer;

KM viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KM immunological disorder; Werner syndrome; developmental disorder;
KM psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
KM neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
KM myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KM angiodysplasia; congenital heart disease; HDR syndrome; gene therapy;
KM polyglutamine disorder; solid tumour.
XX
XX Unidentified.
XX
XX WO200192565-A2.
XX
XX 06-DEC-2001.
XX
XX 06-APR-2001; 2001WO-EP003973.
XX
XX 06-APR-2000; 2000DE-01019058.
XX 07-APR-2000; 2000DE-01019173.
XX 30-JUN-2000; 2000DE-01032529.
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIC-) EPIDENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-090046/12.
XX
XX New nucleic acids or oligomers, useful for diagnosing or treating
PT diseases associated with DNA transcription, e.g. immunological disorders,
PT Werner syndrome, psoriasis, myocardial infarction, solid tumors or
PT cancer.
XX
XX Claim 1; SEQ ID NO 163; 32bp; English.
XX
XX The invention relates to a nucleic acid, which comprises a segment of the
CC chemically pretreated DNA of genes associated with DNA transcription from
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC to the chemically pretreated DNA of genes associated with DNA
CC transcription. The set of oligomer probes are useful for detecting the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC diagnosing or treating the diseases associated with DNA transcription
CC (particularly with the methylation status) e.g. adenosine deaminase
CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner's syndrome,
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC neurological disorders, neurodegenerative disorders, Waardenburg
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiodysplasia, erythropoiesis, congenital heart
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumors
CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC associated genomic DNA molecules of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification but
CC was obtained in electronic format directly from the European Patent
CC Office
XX
SQ Sequence 6874 BP; 1589 A; 259 C; 1935 G; 3091 T; 0 U; 0 Other;
Query Match 4.4%; Score 42.4; DB 6; Length 6874;
Best Local Similarity 50.5%; Pred. No. 1.4;
Matches 103; Conservative 0; Mismatches 101; Indels 0; Gaps 0;
QY 747 TCGAAGACATGATTAATTAATCAACAAATCTACAGCTGGAAGTGGAAATCTGCA 806
DB 2842 TACATTATATTAATTAATTAACCCCTACAAATTCATTTTAAATCAAAAAATTAAT 2783
QY 807 GCAAGCACTGCTTATTAATTAATCTAATGTAATTCATTTTCTTACTTTC 866
DB 2782 ATTATCAATTTCTCATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2723
QY 867 TGAATCTGATCAATGCTCTTATTTCAAAATCTCACTTTAAATCAATTAATTAATTA 926

Db 2722 CTTACTATAATCACAATCCAAAACTTAAAAATATAATATATACAATAA 2663
 QY 927 TGGGTTTATTCAAATACATCATCT 950
 Db 2662 TCAATAAAAATATAATATATATTTT 2639

RESULT 12
 AAS61333/c
 ID AAS61333 standard; DNA; 6874 BP.
 XX AAS61333;
 XX
 DT 29-JAN-2002 (first entry)
 DE Human gene regulation-associated gene oligonucleotide #288.
 XX
 KW Human; Gene regulation-associated gene; severe combined immunodeficiency;
 KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
 KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
 KW renal disease; Preecclampsia; cardiac allograft vascular disease;
 KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
 KW immunostimulant; cardiac; antiinflammatory; coagulant; antiasthmatic;
 KW nephroretropic; gynecological; anti-tumour; immunosuppressive; cytostatic.

OS Homo sapiens.
 XX
 PN WO200177375-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 06-APR-2001; 2001WO-EP003968.
 XX
 PR 06-APR-2000; 2000DE-01019058.
 XX
 PR 07-APR-2000; 2000DE-01019173.
 XX
 PR 30-JUN-2000; 2000DE-01032529.
 XX
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2002-017470/02.
 XX
 PT New nucleic acid sequences from chemically modified genes associated with
 PT gene regulation, useful for analyzing cytosine methylations for diagnosis
 PT and therapy of diseases e.g. severe combined immunodeficiency disease.
 XX
 PS Disclosure; SEQ ID NO 295; 26pp; English.

CC The invention relates to 224 nucleic acid sequences comprising at least
 CC 18 bases of a chemically pretreated gene associated with gene regulation
 CC selected from 43 known genes (or complementary sequences). The chemical
 CC pretreatment converts cytosine bases unmethylated at the 5-position to
 CC uracil or another base with hybridisation behaviour dissimilar to
 CC cytosine, to enable analysis of cytosine methylations. The DNA sequences,
 CC oligomers (or sets/arrays) and method are useful in the diagnosis of
 CC diseases (or predisposition to diseases) associated with gene regulation
 CC and in therapy of such diseases, by enabling analysis of the cytosine
 CC methylation patterns of such genes, kits are provided. They are
 CC especially useful in diagnosis and therapy of e.g. severe combined
 CC immunodeficiency disease, cardiac disorders, haemophilia, solid tumours
 CC and cancer, Werner syndrome, asthma, HDR syndrome, Saethre-Chotzen
 CC syndrome, renal disease, preecclampsia, graft versus-host disease. The
 CC present sequence is a sequence included in the sequence data for this
 CC specification and is associated with the human gene regulation-associated
 CC genes. Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

CC Sequence 6874 BP; 1589 A; 259 C; 1935 G; 3091 T; 0 U; 0 Other;
 SQ

Query Match 4.4%; Score 42.4; DB 6; Length 6874;

Best Local Similarity 50.5%; Pred. No. 1.4;
 Matches 103; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 747 TCGAAGAAGCATTCAGATTATCAACAAATCTCCAGCTGGAGATCGAAACTGCGA 806
 Db 2842 TACATTATATATTAATTAACCCCTACAAATTCACATTTTATTAATCAAAAATATAAAT 2783
 QY 807 GCAAGCAGCTGGCTCTATTAAATTTCTAACATGTAATTTCAATTATCTCTTACTTTC 866
 Db 2782 ATTATCAATTTTCATATATTCATATATATATATTAATTAACAATATTAAGTACTTT 2723
 QY 867 TGAATCTTGCATACATGATGCTCTTATTTCAAAAATCTCAGTTTAAATCATATTAATA 926
 Db 2722 CTTACTATAATCACAATCCAAAACTTAAAAATATAATATATATACAATAA 2663
 QY 927 TGGGTTTATTCAAATACATCATCT 950
 Db 2662 TCAATAAAAATATAATATATATTTT 2639

RESULT 13
 ABN80230/c
 ID ABN80230 standard; DNA; 6874 BP.
 XX ABN80230;
 XX
 DT 15-JUL-2002 (first entry)
 DE Human chemically modified disease associated gene SEQ ID NO 247.
 XX
 KW Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
 KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;
 KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;
 KW antidiabetic; cytostatic; anticonvulsant; ds.

OS Homo sapiens.
 XX
 OS Synthetic.
 XX
 PN WO200200927-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP007536.
 XX
 PR 30-JUN-2000; 2000DE-01032529.
 XX
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2002-130908/17.
 XX
 PT Novel nucleic acid useful for diagnosis and therapy of diseases
 PT associated with development genes such as diabetes, comprises a sequence
 PT of a segment of chemically pretreated DNA of genes associated with
 PT development.
 XX
 PS Claim 1; SEQ ID NO 247; 27pp; English.

CC The invention relates to a nucleic acid (I) comprising a sequence at
 CC least 18 bases in length of a segment of chemically pretreated DNA (II)
 CC of genes associated with development selected from 87 genes listed in the
 CC specification such as ACCPN, ABPN, or ABP1 and comprising one of 350
 CC sequences (ABN79984-ABN8033) or their complements. The invention is
 CC useful for the diagnosis or therapy of diseases associated with
 CC development genes, in particular disease related to homeobox containing
 CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes
 CC associated with congenital heart disease, epilepsy, diseases related to
 CC histone deacetylation, Curranio syndrome, diseases related with the
 CC development of the brain and limb girdle muscular dystrophy and dwarfism.
 CC Oligomers specific to each of the genes are useful for detecting the
 CC methylation state of all CpG dinucleotides within the 350 sequences or

CC (ii) and their complementary sequences, as primer oligonucleotides for
CC the amplification of the 350 sequences, (ii) and/or their complements and
CC as oligomer probes for detecting the cytosine methylation state and/or
CC single nucleotide polymorphisms (SNPs). Note: The sequence data for this
CC patent did not form part of the printed specification but is based on
CC sequence information supplied to Derwent by the European Patent Office
XX
SQ Sequence 6874 BP, 1589 A; 259 C; 1935 G; 3091 T; 0 U; 0 Other;
Query Match 4.4%; Score 42.4; DB 6; Length 6874;
Best Local Similarity 50.5%; Pred. No. 1.4;
Matches 103; Conservative 0; Mismatches 101; Indels 0; Gaps 0;
Qy 747 TGCAGAGACATACATTAATCAACAAATATACAGAGCTGAGAGCGAAGCTGCGA 806
Db 2842 TACATTTATATTAATTAATTAACCCCTACAAATATACATTTATTAATCAAAAAATTAAT 2783
Qy 807 GCAGACCTGGCTCTTATTTAATATTTCTAACATGTAATTTCAATTTATCTTACTTTC 866
Db 2782 ATATATCAATTTCTCATATTTCAATATATATATATATTAATTAACATATTAACATTT 2723
Qy 867 TGATCTTGCTATCAATGCTCTTATTTCAAAAATCTCATTAAATTCATATTAATA 926
Db 2722 CTTACTATTAATCAATCAATCCAAAACCTTAATAAATACTAATAAATATATACAAATAA 2663
Qy 927 TGGGTTATTCATATACATCATCT 950
Db 2662 TCAATTAATATATATATATTTT 2639
RESULT 14
ADP84158/c
ID ADP84158 standard; DNA; 129017 BP.
XX
AC ADP84158;
XX
DT 23-SEP-2004 (first entry)
XX
DE Human ASTR-1 locus DNA representing part of the GPRA gene Segid 1.
XX
XX de; AST-1; asthma; Igg mediated disease; human; chromosome 7p14-p15;
KM GPRA; G-protein coupled receptor for asthma susceptibility; AAA1;
KM asthma associated alternatively spliced gene 1;
KM single nucleotide polymorphism; SNP;
KM chronic obstructive pulmonary disease; cancer; rhinitis; dermatitis;
KM cytostatic; antiasthmatic; transgenic; asthma locus-1.
XX
XX Homo sapiens.
XX
FH Key location/Qualifiers
FT variation replace(1,T)
FT /*tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT replace(93,G)
FT /*tag= b
FT /standard_name= "Single nucleotide polymorphism"
FT replace(184,.185,TAAGATTA)
FT /*tag= c
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "Insertion polymorphism"
FT replace(918,A)
FT /*tag= d
FT /standard_name= "Single nucleotide polymorphism"
FT replace(983,T)
FT /*tag= e
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(987,C)
FT /*tag= f
FT /standard_name= "Single nucleotide polymorphism"
FT variation replace(1201,.1207,CT)
FT /*tag= g
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "Deletion polymorphism"

FT variation replace(1542,T)
FT /*tag= h
FT /standard_name= "Single nucleotide polymorphism"
FT replace(1710,G)
FT /*tag= i
FT /standard_name= "Single nucleotide polymorphism"
FT replace(1818,T)
FT /*tag= j
FT /standard_name= "Single nucleotide polymorphism"
FT replace(1827,T)
FT /*tag= k
FT /standard_name= "Single nucleotide polymorphism"
FT replace(2254,C)
FT /*tag= l
FT /standard_name= "Single nucleotide polymorphism"
FT replace(2537,A)
FT /*tag= m
FT /standard_name= "Single nucleotide polymorphism"
FT replace(3877,C)
FT /*tag= n
FT /standard_name= "Single nucleotide polymorphism"
FT replace(4012,A)
FT /*tag= o
FT /standard_name= "Single nucleotide polymorphism"
FT replace(4631,C)
FT /*tag= p
FT /standard_name= "Single nucleotide polymorphism"
FT replace(4689,G)
FT /*tag= q
FT /standard_name= "Single nucleotide polymorphism"
FT replace(4961,G)
FT /*tag= r
FT /standard_name= "Single nucleotide polymorphism"
FT replace(5442,G)
FT /*tag= s
FT /standard_name= "Single nucleotide polymorphism"
FT replace(5634,A)
FT /*tag= t
FT /standard_name= "Single nucleotide polymorphism"
FT replace(5850,G)
FT /*tag= u
FT /standard_name= "Single nucleotide polymorphism"
FT replace(6312,T)
FT /*tag= v
FT /standard_name= "Single nucleotide polymorphism"
FT replace(6392,T)
FT /*tag= w
FT /standard_name= "Single nucleotide polymorphism"
FT replace(6485,A)
FT /*tag= x
FT /standard_name= "Single nucleotide polymorphism"
FT replace(6522,G)
FT /*tag= y
FT /standard_name= "Single nucleotide polymorphism"
FT replace(6646,G)
FT /*tag= z
FT /standard_name= "Single nucleotide polymorphism"
FT replace(6739,A)
FT /*tag= aa
FT /standard_name= "Single nucleotide polymorphism"
FT replace(6760,C)
FT /*tag= ab
FT /standard_name= "Single nucleotide polymorphism"
FT replace(6786,.6817,(TAAAT))
FT /*tag= ac
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "Deletion polymorphism (TAAAT)8 to (TAAAT)7"
FT replace(6820,.6822,AT)
FT /*tag= ad
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "Deletion polymorphism"
FT replace(7125,A)
FT /*tag= ae

CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 CC sequences (AB101840-AB116175) and the encoded proteins (AB85737-
 CC AB872072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX

Sequence 15320 BP; 4914 A; 3114 C; 3197 G; 4095 T; 0 U; 0 Other;

Query Match 4.3%; Score 41.2; DB 4; Length 15320;

Best Local Similarity 57.9%; Pred. No. 4.2; Mismatches 53; Indels 0; Gaps 0;

Matches 73; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy	819	TCTTATTAAATATTCTAACATGTAAATTCATTATCTTACTTCTGATCTTGCTAT	878
Db	7744	TATTATTCAGAACTACTAGAGAGTATTATATCACTACTTCTTATATTGTTAG	7803
Qy	879	CACATGTCTCTTATTTCAAAAATCTCACTTAAATTCATATAAATGAGTTATTCA	938
Db	7804	AACATGCTTATCATACATAGACTATGCTCTTAAATATATAATATATTATTTTGTAA	7863
Qy	939	AATACA	944
Db	7864	AATAAA	7869

Search completed: April 9, 2005, 04:02:54
 Job time : 618 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 9, 2005, 01:32:09 ; Search time 4370 Seconds
(without alignments)
10578.095 Million cell updates/sec

Title: US-09-087-136-2
Perfect score: 954
Sequence: 1 atgcacgaatagatccacc.....ttcaatracatccttgcac 954

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 2422767955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*

- 1: gb_ba:*
- 2: gb_hcg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	439.2	46.0	32981	3 U00047	U00047 Caenorhabditis
C 2	51.8	5.4	165933	1 AP004174	AP004174 Mycoplasma
C 3	51.2	5.4	1141	6 AX083744	AX083744 Sequence
C 4	47.6	5.0	2000	6 AX655393	AX655393 Sequence
C 5	46.8	4.9	260699	2 AC006893	AC006893 Caenorhab
C 6	46.2	4.8	254449	3 AE014817	AE014817 Plasmodiu
C 7	45.8	4.8	123905	9 AC116311	AC116311 Homo sapi
C 8	45.8	4.8	154535	9 AC114283	AC114283 Homo sapi
C 9	45.8	4.8	157165	2 AC104112	AC104112 Homo sapi
C 10	45.8	4.8	180797	2 AC116338	AC116338 Homo sapi
C 11	45.6	4.8	199421	2 AC051636	AC051636 Homo sapi
C 12	45.6	4.8	207905	10 AL672013	AL672013 Mouse DNA
C 13	45.2	4.7	38692	3 AC116919	AC116919 Dictyoste
C 14	45.2	4.7	165062	9 AC087834	AC087834 Pan trogl
C 15	45.2	4.7	178824	9 AC146421	AC146421 Pan trogl
C 16	45.2	4.7	192929	2 AC005505	AC005505 Plasmodiu
C 17	45.2	4.7	250713	3 AE014850	AE014850 Plasmodiu
C 18	45.2	4.7	256172	2 AC005139	AC005139 Plasmodiu
C 19	45.2	4.7	314016	2 AC146012	AC146012 Pan trogl

C 20	45	4.7	1813	9 COL11A1G08	AF101086 Homo sapi
C 21	44.6	4.7	6291	6 AX344478	AX344478 Sequence
C 22	44.6	4.7	6291	6 AX344640	AX344640 Sequence
C 23	44.6	4.7	6292	6 AX251553	AX251553 Sequence
C 24	44.6	4.7	238626	2 AC108529	AC108529 Rattus no
C 25	44.4	4.7	8346	6 AX323713	AX323713 Sequence
C 26	44.4	4.7	348313	1 CR378666	CR378666 Photobact
C 27	44	4.6	138316	2 CR457441	CR457441 Dario rer
C 28	44	4.6	212866	5 BX855612	BX855612 Zebrafish
C 29	43.4	4.5	471	3 AF176058	AF176058 Eubazus b
C 30	43.4	4.5	67800	9 AC006981	AC006981 Homo sapi
C 31	43.4	4.5	276034	1 AE017015	AE017015 Bacillus
C 32	43.2	4.5	123227	2 AC151239	AC151239 Bos tauru
C 33	43.2	4.5	162352	2 AC150523	AC150523 Bos tauru
C 34	43.2	4.5	232226	2 AC110915	AC110915 Mus muscu
C 35	42.8	4.5	158390	2 AP004633	AP004633 Oryza sat
C 36	42.8	4.5	158444	8 AP005822	AP005822 Oryza sat
C 37	42.6	4.5	168042	2 AC146411	AC146411 Pan trogl
C 38	42.6	4.5	50000	6 AX392733	AX392733 Sequence
C 39	42.6	4.5	232392	14 AF250284	AF250284 Amsacta m
C 40	42.4	4.4	6874	6 AX252034	AX252034 Sequence
C 41	42.4	4.4	6874	6 AX271984	AX271984 Sequence
C 42	42.4	4.4	6874	6 AX323675	AX323675 Sequence
C 43	42.4	4.4	6874	6 AX344822	AX344822 Sequence
C 44	42.4	4.4	6874	6 AX348813	AX348813 Sequence
C 45	42.4	4.4	147414	2 AC078779	AC078779 Homo sapi

ALIGNMENTS

RESULT 1	U00047/c	32981 bp	DNA	linear	INV 11-SEP-2004
LOCUS	U00047	Caenorhabditis elegans	cosmid ZK418	complete sequence.	
DEFINITION	U00047	U00047.1	GI:470368		
ACCESSION	U00047				
VERSION	U00047.1				
KEYWORDS	HTG.				
SOURCE	Caenorhabditis elegans				
ORGANISM	Caenorhabditis elegans				
	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.				
REFERENCE	1 (bases 1 to 32981)				
AUTHORS	WormBase Consortium				
CONSTRM	Genome sequence of the nematode C. elegans: a platform for				
TITLE	investigating biology. The C. elegans Sequencing Consortium				
JOURNAL	Science 282 (5396), 2012-2018 (1998)				
MEDLINE	99069613				
PUBMED	9851916				
REFERENCE	2 (bases 1 to 32981)				
AUTHORS	Fulton, L.				
TITLE	The sequence of C. elegans cosmid ZK418				
JOURNAL	Unpublished (2001)				
AUTHORS	3 (bases 1 to 32981)				
REFERENCE	Waterston, R.				
AUTHORS	Direct Submission				
TITLE	Submitted (08-APR-1994) Department of Genetics, Washington				
JOURNAL	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.				
	Louis, MO 63110, USA				
	4 (bases 1 to 32981)				
	Waterston, R.				
REFERENCE	Direct Submission				
AUTHORS	Submitted (28-JUN-2001) Department of Genetics, Washington				
TITLE	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.				
JOURNAL	Louis, MO 63110, USA				
	5 (bases 1 to 32981)				
REFERENCE	Waterston, R.				
AUTHORS	Direct Submission				
TITLE	Submitted (29-MAY-2002) Department of Genetics, Washington				
JOURNAL	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.				
	Louis, MO 63110, USA				
	6 (bases 1 to 32981)				

AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (22-NOV-2002) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
REFERENCE 7 (bases 1 to 32981)
AUTHORS Wilson, R.
TITLE Direct Submission
JOURNAL Submitted (15-JUN-2003) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
REFERENCE 8 (bases 1 to 32981)
AUTHORS Wormbase Consortium
TITLE Direct Submission
JOURNAL Submitted (11-SEP-2004) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
COMMENT Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
email: submissions@waterston.wustl.edu and jee@sanger.ac.uk
NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections or, or longer because we provide a small overlap between neighboring submissions.
This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m3 subclone.
For a graphical representation of this clone sequence and its analysis see:
http://www.wormbase.org/db/seq/sequence?name=ZK418;class=Sequence
NEIGHBORING CLONE INFORMATION
The 5' clone is R01H2, 200 bp overlap; the 3' clone is B0280, 200 bp overlap. Actual start of this clone is at base position 197 of ZK418; actual end is at 7055 of B0280.
NOTES:
Coding sequences below are the result of integration and manual review of the following data: computer analysis using the program GeneFinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yoji Kohara (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C. elegans ORFome cloning project (http://worfcd.bfci.harvard.edu/), similarity to other proteins from Blastx analyses (http://blast.wustl.edu/), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans Genbank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).
Location/Qualifiers
1..32981
/organism="Caenorhabditis elegans"
/mol_type="genomic DNA"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="III"
/clone="ZK418"
92..1634
/gene="ZK418.3"

CDS
/locus_tag="ZK418.3"
join(92..178,1121..1234,1283..1405,1473..1634)
/gene="ZK418.3"
/locus_tag="ZK418.3"
/standard_name="ZK418.3"
/note="contains similarity to Mus musculus Transmembrane protein 17; TR:O8K0U3; coded for by the following C. elegans cDNAs: YK686e01.3"
/codon_start=1
/product="Hypothetical protein ZK418.3"
/protein_id="AA50687.1"
/db_xref="GI:470371"
/db_xref="WormBase:ZK418.3"
/translation="MADPSCINVCIPNCSGLQNLGCCPRRLHSLLSTLYQIYARSA CLVHLGSEVRLGIGFYGNLAENNSALFGLITITIIIOPLTPVLANSGFLNPLEY VFYMTITVFALMLMKTAEREVAKYINKIOQERVADELKTSKSPKSGRTIKQOLS R"
complement(1751..3301)
/gene="lin-37"
/locus_tag="ZK418.4"
complement(join(1751..1868,2059..2228,2276..2555,2993..3127,3177..3301))
/gene="lin-37"
/locus_tag="ZK418.4"
/standard_name="ZK418.4"
/note="contains similarity to Arabidopsis thaliana Hypothetical protein Atg32840; TR:O8G1K6; coded for by the following C. elegans cDNAs: CB102055, CB102059, CB102062, CB102064, CB102065, CB102066, CB102067, CB102169, CB102170, CB102172, CB102173, CB102175, CB102177, OSTRF038D2_1, YK71d12.5, YK90a5.5, YK96b5.3, YK96b5.5, YK191f6.3, YK550h7.3, YK550h7.5, YK560h10.3, YK560h10.5, YK686e01.3"
/codon_start=1
/product="Abnormal cell lineage protein 37"
/protein_id="AA50688.2"
/db_xref="GI:14574553"
/db_xref="WormBase:ZK418.4"
/translation="MSEIIPPLAEFLPEGGDRNARQNDPLISGSPPLSPSPRKLTSLSIDPVPSPDPMKFARKLGNLITTKHPSEITIGVLPEDYTRADEPFGQRPGRPRMPRESSTLSMESPRKTMDSKIMELRKPPEMAGREESYSLGRMVKGMNMEYEPKQARDYAPNLAVDYLACREIRHMPDPKSIDPLPVEPSRIDEPATVDPRYETDLKNEYIRHMKYQKGMCAHORRTAPHARSIALINKIQPGSEKTVQALGLI"
complement(3301..3302)
/note="SL2 trans-splice acceptor; see YK90a5.5"
complement(3582..4582)
/gene="ZK418.5"
/locus_tag="ZK418.5"
complement(join(3582..3717,3768..3904,3955..4152,4330..4463,4510..4582))
/gene="ZK418.5"
/locus_tag="ZK418.5"
/standard_name="ZK418.5"
/note="contains similarity to Mus musculus Mus musculus adult male small intestine cDNA, RIKEN full-length-enriched library, clone:2010004B11
product:SEVEN TRANSMEMBRANE DOMAINS PROTEIN homolog; TR:Q9P8P0; coded for by the following C. elegans cDNAs: OSTRF037F8_1, OSTR037F8_1, YK329h10.3, YK329h10.5, YK600e3.3, YK600e3.5"
/codon_start=1
/product="Hypothetical protein ZK418.5"
/protein_id="AA50689.2"
/db_xref="GI:14574554"
/db_xref="WormBase:ZK418.5"
/translation="MSFPHFNCFALAPAPFYVYKSGINYSIMKCATASGGYLTLQALKLIIATFPFALDSEGSIVPEFLKSSADIIDVIGLHLMTNFLAGKGEVRFVCGIAGGFASVARIVLVWVGARGATFWRWGTIDSSADLVIVSLCLTWMITRTPKKFLVSPILAMCVSTFYVQVGHFSLYGMSLAPPAVSIAVAIILVVVYSNRTASTRKNE"
complement(4600..4601)
/note="S11 trans-splice leader [TEC-RBD]; see ce1031703ab0112"

misc_feature

gene

gene	complement(5033..9583)			
	/gene="ZK418.6"			
	/locus_tag="ZK418.6"			
	complement(5033..5578,5882..6128,6423..6777,			
	6943..7091,7366..7533,7789,8102..8318,8311..8501,			
	8553..8718,8764..8897,9314..9400,9512..9583))			
	/locus="ZK418.6"			
	/locus_tag="ZK418.6"			
	/standard_name="ZK418.6"			
	/note="contains similarity to Pfam domain PF00001 (7			
	transmembrane receptor (rhodopsin family); coded for by			
	the following C. elegans cDNAs: OSTR050H5_1"			
	/codon_start=1			
	/product="Hypothetical protein ZK418.6"			
Query Match	46.0%;	Score 439.2;	DB 3;	Length 32981;
Best Local Similarity	74.4%;	Pred. No. 7.2e-98;		
Matches 698;	Conservative 0;	Mismatches 3;	Indels 237;	Gaps 2;
Qy	254	ATTATATCTCGTCGATGAGAGCCCGGCGCAAGAGCGTCCACGAGTGGCCCTGCTA	313	
Db	2562	ATTTCAGTCGTGTCGATGAGAGCCCGGCGCAAGAGCGTCCACGAGTGGCCCTGCTA	2503	
Qy	314	AGATGCCGCTCAGCAATCTTCACTTATGGAATCACCAGCAAGACTATGACTC	373	
Db	2502	AGATGCCGCTCAGCAATCTTCACTTATGGAATCACCAGCAAGACTATGACTC	2443	
Qy	374	GTGATCTTAAATATATATGTTGTAATGCGTGAAGAACATTCGAAATGATAGCTGACGTT	433	
Db	2442	GTGATCTTAAATATATATGTTGTAATGCGTGAAGAACATTCGAAATGATAGCTGACGTT	2383	
Qy	434	TTGAGAGAAATATATCTTGTGAGAGATGGTTAAAGACACATGTAATTAATATG	493	
Db	2382	TTGAGAGAAATATATCTTGTGAGAGATGGTTAAAGACACATGTAATTAATATG	2333	
Qy	494	AACCAATTAAGACTCAAGAGACAGACTATGACCGAATCTGGCTGTT-----	540	
Db	2322	AACCAATTAAGACTCAAGAGACAGACTATGACCGAATCTGGCTGTTGGTGAAG	2263	
Qy	541	-----GATTATCTTGATGTCGCGAGATTCA	566	
Db	2262	ATTAAATGAAAAATGTTATTAATTTCTTCCTGAGGATTAATCTTGATGTCGCGAGATTCA	2203	
Qy	567	TGCAATGCGAGTCCAGTAAATCAATTCCTGAGTGGCAATTTGTCATCTAAGATGCA	626	
Db	2202	TGCAATGCGAGTCCAGTAAATCAATTCCTGAGTGGCAATTTGTCATCTAAGATGCA	2143	
Qy	627	TGAATTCGACGCTCAGTCGATCCAAAGATATGAAACAGATTGAAAAATGAATACATTG	686	
Db	2142	TGAATTCGACGCTCAGTCGATCCAAAGATATGAAACAGATTGAAAAATGAATACATTG	2083	
Qy	687	TCATTTGAAAAACAAGTCAAAAAAG-----	709	
Db	2082	TCATTTGAAAAACAAGTCAAAAAAGGTTAACCTTCTTGCCATTGAAAAATTCGTTAAAA	2023	
Qy	710	-----	709	
Db	2022	ATTATATTTCTTAAAGTATTAACGTAACGTGCTCTATGTAGAAATTCATTTGCGAATT	1963	
Qy	710	-----	709	
Db	1962	TTGTATCAAACTATTAATAATTAATATATGTAACGATCTGTCATCTACTATTTCT	1903	
Qy	710	-----GTTGGTGTGCTTCATCAACGTCGTCGGA	736	
Db	1902	TCAAGTCAGAGAAATCAAAAAAAATTTCAATTTGTTGTCATCAACGTCGTCGGA	1843	
Qy	737	CTGCTCCCAAGCAAGAGACTAGATTAATCAACAAATCTACAGCTGAGAGTGA	796	
Db	1842	CTGCTCCCAAGCAAGAGACTAGATTAATCAACAAATCTACAGCTGAGAGTGA	1783	
Qy	797	AAACGTGAGACGACACTTGTCTTATTAATATCTTAACATGTAATTCATTTATTC	856	

Db	1782	AAACGTGAGACGACACTTGTCTTATTAATTAATTTCTAACATGTAATTCATTTATTC	1723
Qy	857	TCTTACTTTCGATCTGATCATCATGCTCTTATTTCAAAATTCACCTTAAATTC	916
Db	1722	TCTTACTTTCGATCTGATCATCATGCTCTTATTTCAAAATTCACCTTAAATTC	1663
Qy	917	ATATTAATTAATGGGTTTATTCATATCATCATCTTGCAC	954
Db	1662	ATATTAATTAATGGGTTTATTCATATCATCATCTTGCAC	1625
RESULT 2			
AP004174/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Mycoplasma penetrans HF-2			
Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
P			

```

/codon_start=1
/evidence=not experimental
/transl_table=4
/product="conserved hypothetical protein"
/protein_id="BAC44679.1"
/db_xref="GI:26454351"
/translation="MSYLILYMOKKRCSVSNDKVINYTWLSRKINYSANVSN
YEIVVTRPLHVKSSVDVDFVSKYSPWKRILDKNNYFENLKNPFRILGHXYDLV
QLGNSKSTYKLIENNIVYLNKINIEKKEVYIRLITKKEIYEAARHKANNIILQVN
KRDVDMKRAMAYTHNKKKEIYFSTKLVVGFPIIDVYICHELAAHLIFPHMSDFML
VSKLCPNKKKIKVLSKF"
/complement(2233. .4221)
/gene="MYPE8890"
/complement(2233. .4221)
/gene="MYPE8890"
/standard_name="uvrB"
/codon_start=1
/evidence=not experimental
/transl_table=4
/product="excinuclease ABC subunit B"
/protein_id="BAC44680.1"
/db_xref="GI:26454352"
/translation="MNTNKNVDFDKSKFSGDQOALIEKIVGKEDKFKSVOYLVG
ATGKGTPTMANITQSLNCKTLVLAHNKTLAQLYSEFKMFPDNKVEYFVAPFYQ
PEARLPKNTYIEKNSQSNDDIEMNLNLSANLSIERDDVYVAVAAIYASPPDPF
KNRYAVKKNQVPIKQVQYELVNIYQNDVDVTHGRFVKGDLMIIGFPTDERNVR
ISFPGDEIEIAIDPLTKQVITKDEIYLPAGEVYLARDLLESLIKRIRDELTEH
KNPLKONKLEAORILERTNIDIESFKREGFCGIENYSPRLDLNRPDLNFFITDYF
DKMLILVDSHMMVPIQIRGMVNTDRSKRTYVEYGFRLPSALDRPLNFFELSKD
KLVYSATPMEWIEESOOTEOIIRPTGVDTIIRPTQOQILIDIEELKQIANK
ERTPTVMTIRLAEIETIRKNIYVAYLHNEKTERKILNNLRKQIYDVVIGIN
LIREGIDVPEVSVYLIDADKGFSPRSKSLQIGRASRNNGKVMYADITTEAMD
KAIREBERKNIIIEFNKHNITPKTVIRPIALDLSKNDADINIDELSKKKSEKI
DRTVKLREMLAANOEVERAAYIRDIIMELANKK"
/complement(4289. .5104)
/gene="MYPE8900"
/complement(4289. .5104)
/gene="MYPE8900"
/codon_start=1
/evidence=not experimental
/transl_table=4
/product="conserved hypothetical protein"
/protein_id="BAC44681.1"
/db_xref="GI:26454353"
/translation="WIKILFIQDISGENKTAIVSELPKLNKEINIDVYANAMNTTK
GGLMWTDRKLSGIDPFTWGNHTWHRDIYILKTKNNIIRPANKSVFESKYG
RRTKAVTINKKTRITNLGISVSRMDQNPTELQNIIDNSKRTDPHIDHCETT
SRKNAFIFNFGAVALIGTHVQINDAKIINNTAYITDGMGANGVIGAKPDSI
LYMFGYSRFLIEQIGPYQFCGVIISFNHTKKVQIKKVILODYKIKISK"
/complement(5203. .7419)
/gene="MYPE8910"
/complement(5203. .7419)
/gene="MYPE8910"
/codon_start=1
/evidence=not experimental
/transl_table=4
/product="ATP-dependent DNA helicase"
/protein_id="BAC44682.1"
/db_xref="GI:26454354"
/translation="WQKITIDSENEKYEAITAPDVPMLVINGAGSGTAVLYTAV
YLNLNATVSDRIIGFTFNKAANEMKORTTKVYPNNSFKYIGTSPSVCGLIRBETIK
ELNIGKINSNFTIIDEDDONSILKEIYKLNPFKOLINNYKVLCSWISKISKNIYAN
DYVEMLSFLDNTYDGVKDIIPAVYEGYINPFRDNLDPDOLIKYARVYIEIEIR
TKMONRPFYILVDEFDOTNYDOYQIILSKNKNIRPAGDDPQMTYSRGAQENIFE
VFKDFSDSLTIVLEKRYSTKNLINSITLONKNIIRKNIYKNIYDSTGESIIFYEA
KPYRKAVKQDIATMKLIIVSDVSILTRIVTPRQISATYFKIKNYAVANNISIPA
AFKRVDEIEGIQRTKNAVQAFYDSMIEIEIKKYGSIEMIDIVKTYVMIMEDG
BEHRTENIIEBKNAKNEQNGNSNINFLQIALTALLTNDSEENKQNDQSVLMTV
HASKGLFKIVLFIQENIIFPSIRLEENIDEBERIVAVITAKENULYISCGNGT
SFGNREMEREPSEIFSEIGDALIERKILTKTKTKRATQGNNSKSVHSPFHE
SEKTYVYGGELIHDKREGIVAVNSGNIIDVSKFNDYGRKSLFKQHVSIRSLNOK"
/complement(7422. .8219)

```

```

CDS
/gene="MYPE8920"
/complement(7422. .8219)
/gene="MYPE8920"
/codon_start=1
/evidence=not experimental
/transl_table=4
/product="conserved hypothetical protein"
/protein_id="BAC44683.1"
/db_xref="GI:26454355"
/translation="MDYNLDKNOKKKYNLASALRSLTRFLDLIVFLILVAFPAIF
SNVLYNFSSTNNPDLWYSIDSMRIEFLTFEVFVIPLSYFVFPVPAOGTYLESKIEK
IRIVSTSLKINENRKIFKNHFLFLVOLRELTLTILISFALLIGTVALPEKDY
IDVFNQTRSDPTLNGSTNNVAVIYPOAFCTIGILNTVILVYACTSRKRSFTHDIS
TVVKNVETYSDEKNNNLNKNKKOPVYKYNLNGEINPDEIFPKETN"
/complement(8233. .9192)
/gene="MYPE8930"
/complement(8233. .9192)
/gene="MYPE8930"
/codon_start=1
/evidence=not experimental
/transl_table=4
/product="glycosyltransferase"
/protein_id="BAC44684.1"
/db_xref="GI:26454356"
/translation="MKITFIYHLYNNSVNLKSKLSLEQTSNPFVPIIDPASDEC
KNILNFRDLSDKKFKIVLRFENYGRFSYNLGLERASGYYVFAEHSNLSFPDVEBI
INTFKKQDFIIFKTNVNLNDIDSNDELPMAYPAKSLIDFRKSENGVRYTNLDIL
KFLPKKKLFIYKYNLPLYLPDVLVYMAIYPAKSLIDFRKSENGVRYTNLDIL
EAAFIISFKTSQSLNNIKNKFPQWLPGLCTIEDPLIMPDSYSEKILSIAITKWE
TLEKIDVSRKRSKSNLNLNTHLKQYIKGFPTYSFVKNNLIS"
/complement(9298. .9600)
/gene="MYPE8940"
/complement(9298. .9600)
/gene="MYPE8940"
/codon_start=1
/evidence=not experimental
/transl_table=4
/product="hypothetical protein"
/protein_id="BAC44685.1"
/db_xref="GI:26454357"
/translation="MKQVDRKONNDKKKSSFPDFLTHKTDIDNOFGKQESKSKRN
QIILICITVIFIRICVAVGVILWTKRKLTSIQQIDYIDVDTSSNVSFKVSNLT"
/complement(9626. .10978)
/gene="MYPE8950"
/complement(9626. .10978)
/gene="MYPE8950"
/codon_start=1
/evidence=not experimental
/transl_table=4
/product="cytosol aminopeptidase"
/protein_id="BAC44686.1"
/db_xref="GI:26454358"
/translation="MENIDRKOITFKALDIRSEINLDSKAYTEPEKENIVYFLVDKN
KDLPRIDIVYKATENFKRIDIDLFLFLEVGLSPSEIIDLITVIVAFRPNKI
PFTMKQVSGSKLYTVNEKYKOVITNALVINDAOYFCGNLDRSSQIYPSPIEB
VKLFPADVOGYSISVNLKEDIKREKNNILVINDAGSIRPRICITBYIANPSNGK
AYVKGITDYSGNNITKMTYMKRMKFDMSGAAVAWATYALAKNNIRINVGIGMT
ENLPGLAIRBDIIVSHSKTYEIDNTDSEGLVLADGLSAVKKYNAIKYVADIATL
TGAMVPSLGVYISGVWATINDMKVPEKTAADKSGEYVWRIPLHEDYRKSIDRSIDIA
NNSNDRASSSSAACFLVEFTGVDVYIHLVDVALLTVDVKNMGQTIKTLVNFATQI
"
/complement(11139. .11846)
/gene="MYPE8960"
/complement(11139. .11846)
/gene="MYPE8960"
/standard_name="grpE"
/codon_start=1

```

Query Match 5.4%; Score 51.8; DB 1; Length 165933;
 Best local similarity 56.0%; Pred. No. 0.11; Indels 0; Gaps 0;
 Matches 98; Conservative 0; Mismatches 77;

774 AATCTACCAAGCTGAGAGTGGAAAACGTGAGGACGACCTTGCTTATTATAATATT 833

gene

Kye, S., Chan, M.-S., Nene, V., Shallom, S. J., Sub, B., Peterson, J., Anjoui, S., Perce, M., Allen, J., Selengut, J., Haft, D., Mather, M. W., Vaidya, A. B., Martin, D. M. A., Fairclamb, A. H., Fraunholz, M. J., Roos, D. S., Ralph, S. A., McFadden, G. I., Cummings, J. M., Subramanian, G. M., Mungall, C., Venter, J. C., Carucci, D. J., Hoffman, S. L., Newbold, C., Davis, R. W., Fraser, C. M. and Barrett, B.

TITLE
Genome sequence of the human malaria parasite *Plasmodium falciparum*
JOURNAL
Nature 419 (6906), 498-511 (2002)
PubMed
12368864
2 (bases 1 to 254449)
AUTHORS
Gardner, M. J.
TITLE
Direct Submision
JOURNAL
Submitted (13-SEP-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1. .254449
/organism="Plasmodium falciparum 3D7"
/mol_type="genomic DNA"
/isolate="3D7"
/db_xref="taxon:36329"
/chromosome="14"
repeat_region
712..780
/rpt_type=tandem
gene
1083..>2856
/locus_tag="PF14_0066"
/join(<1089..2095,2658..2707,2795..>2856)
/locus_tag="PF14_0066"
/join(1089..2095,2658..2707,2795..2856)
/locus_tag="PF14_0066"
/codon_start=1
/product="hypothetical protein, conserved"
/protein_id="AA036678.1"
/db_xref="GI:23497130"
/translation="MEKSKRYISLKKMERKKFEKYLRLQIMDNIYKGIKLEIKMKNI
IPTEIRRELKNIPIHNILISIKIKELKYDRAVYLFOCKMDEKLEATSDGSHKSLC
ISSQIGCFEGCKFCATGQIGIKRQIDIDETDOLYFOSKVDIKNISPMGGEPLAN
PYVFDSDIOFENDNLFSISNRINISVGLLPGLKLNITPPQVNLAFSLSPTEER
DQLVINKLPFNEVFDLDERIAKTGRWISYILKNLSDKHAEALSDHICKP
NNIRYLNVCLIPYKGNRIYINISFEYIYIYLIIKKILCKYIMHPILYKYGIE
DINFGYSIDAACQLVADYEPKPKKEIKESKNSILL"
repeat_region
2114..2150
/rpt_type=tandem
repeat_region
2423..2460
/rpt_type=tandem
repeat_region
2918..2947
/rpt_type=tandem
repeat_region
2945..2973
/rpt_type=tandem
gene
3572..>7390
/locus_tag="PF14_0067"
/locus_tag="PF14_0067"
/locus_tag="PF14_0067"
/locus_tag="PF14_0067"
/product="hypothetical protein"
/protein_id="AA036679.1"
/db_xref="GI:23497131"
/translation="MIKKDILIFINIIYLILNAYCKEMCKATPEYKSKDVAIGISG
DSIRMIETPIISKVDYLSNVSLSNGYSKTRITIIENEDGLNKITFRSD
IGNPEYIHIKUNSONKMKCKKITVWKDYKWFDCIGSLNDKPESTYFLSGNKIYI
AYVQTKDIEAATGTIEIILGNKRSNTVHLHGSAGGLKKIKFOASVGLLEBDI
ILNNSLNDPMYCDPVKIKSDNNKIYFNVKSWIGAPYDKNVKNIKSDTIEGTSKD
IDCHIRADILINTNNKLQNKYQIFKVRCPONCHSEFALIGSSLSHSPSTICA
AIDHGISPSGGEIIVWYSELNHYITIKETIPHELEALDPSARDEKNPFRFYHD
SIDDIISVRIYDPSFKLSISLGRLEIRLAKNSGTYCVKPNFRFNDAAKAKCXL
GFENGHIKENCANLNGONYCAGKYKYPASSGILCTGHEKQISKCNADDFSHCVDDHD
DVIIQCLHYSNELINDAIRLVINGAPNSNGIGRLQIYHGVGVSCHGWYKAE
NIVCHELGNGLKNGFSHSCDTSIGENLCGPTEKINAVNIKKCGEKLRLRCPSHE
THDIIYCSHDEIVGCVGGBNNGSGEONKXHLINLEKKKHPRKIELTCRDKMLSSS

NSVATGVDYELASCPEKCBDEVGIVKGTFLTYDSPICKSAHSGLVPNNIADIVL
SIATHNRFQTKRNNIIESHDFKTSKSFSTISITMGLREERSNPSDEPMNKNE
IDLTYDHSVFNKRLDHIHTSMKRTFPMIATPGVGRNGKNDITDNLNENEKIR
LSNFTFVYFLNKGDEGWRTLSHCEBGSISINEDNELIIONCPHLKFKPK
KFGQYHISLVFNKINIKNTLYLYNGKVIETKNTYNTFLSGDLIGRNOPTTKQYFIG
NHLVEIYKTVLSECEIKESLNSLSLEYLMENILEQMSNKKDKKKQKRGVQKT
IDGRDCTPCSKNMINDLOINTQOINLKCODNLSQFNGKISGQPLVCEICTK
SKYIKGTNNYPTPDDSSICRAIHAGIYKPMGNKNTFVRIYEGILEYKSGHFG
ILSKESKQSRLSPSVSENEENIFTSCTGQPLNLSVGEKRTINCNSKCNITIKDI
YGINIYSPISVLCRAIHSGLSNQGLVELITVGQDEEFKSTONNVSPSSNNHS
SITPNKMEML"
repeat_region
7408..7434
/rpt_type=tandem
repeat_region
7606..7643
/rpt_type=tandem
repeat_region
8138..8175
/rpt_type=tandem
repeat_region
8318..8348
/rpt_type=tandem
gene
8776..>10018
/locus_tag="PF14_0068"
/join(<8776..8782,9069..>10018)
/locus_tag="PF14_0068"
/join(8776..8782,9069..10018)
/locus_tag="PF14_0068"
/codon_start=1
/product="fibritin, putative"
/protein_id="AA036680.1"
/db_xref="GI:23497132"
/translation="MTDSPRGSQNPFRKNSNPFQKGNNAVRKSKNMKFGNGRGGG
GGRRGG
LKGKSDIIVNTNIPGSEYVGERYVTEDEKILEYVNMNPFSLKALCMGNGNP
IKRGSKTLYIGAAAGTISVSHSDVNGEGVYIVAFESHSRSDITNNSKRSNIVPIY
EDARQPKRYKMLVDIMVVPADVAQPOQARIVANNAIMPLTGGMPFIISKANCVDST
AKEVVPASMEKIKKENCCKPKELTLEPPRHAIYLGMYR"
repeat_region
8872..8910
/rpt_type=tandem
repeat_region
9180..9283
/rpt_type=tandem
repeat_region
9188..9282
/rpt_type=tandem
repeat_region
10250..10291
/rpt_type=tandem
repeat_region
11060..11095
/rpt_type=tandem
gene
11566..>12129
/locus_tag="PF14_0069"
/locus_tag="PF14_0069"
/locus_tag="PF14_0069"
/locus_tag="PF14_0069"
/complement(<11566..>12129)
/locus_tag="PF14_0069"
/complement(11566..12129)
/locus_tag="PF14_0069"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA036681.1"
/db_xref="GI:23497133"
/translation="MAKYKKEENDEGHNGDEPNKIKKLEDKKTKOISPERENIK
KSDPYNDLKLIHINIDETRKCIRESKSHQKHGNMNEYIKRNNIYNMKAVETTF
TKLEDILNKQDKQYNTFKEELDKNTFELKRQMEYCKKLSNYDBIKOTIYVNFND
IKNPKLNSKXKXMFICITYILBRIBV"
repeat_region
12349..12404
/rpt_type=tandem
repeat_region
12512..12547
/rpt_type=tandem
repeat_region
12742..12781
/rpt_type=tandem
repeat_region
12969..13026
/rpt_type=tandem
repeat_region
13821..13862
/rpt_type=tandem
repeat_region
14163..14245
/rpt_type=tandem
gene
14433..>17104
/locus_tag="PF14_0070"

Estimated Total Number of Errors is 0.
Location/Qualifiers
1. 154535
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2319M2"

ORIGIN

Query Match 4.8%; Score 45.8; DB 9; Length 154535;
Best Local Similarity 52.3%; Pred. No. 3.4;
Matches 101; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 761 CATTAATCAACAAATCTACAGCGCTGAGAGTGAGAAACTGTGAGCAACACTTGGTC 820
DB 52130 CATCCAGAAACAAATCTTGGATCTTCAATCCAGTCAAGTAAACACTTAACCAATGAC 52189
QY 821 TTATTTAAATATCTAATCATGTAAATTTCAATTTATCTTCTGATCTTGCATCA 880
DB 52190 ACATTCAAAACCTACGTGCTTCTCTCTCTCAATCTTAATGTTTATGATGCA 52249
QY 881 CATCTCTTATTTCAAAATCTCACTTAAATCATATAATTAATGAGTTTATCAAA 940
DB 52250 TAACTCATATCTTATATAAATATATTATTATTAATGTTTATGATGTTATTTAAAT 52309
QY 941 TACTCATCTTGA 953
DB 52310 AATTTAACTTTTA 52322

RESULT 9

LOCUS AC104112 157165 bp DNA linear HTG 04-DEC-2001
DEFINITION Homo sapiens chromosome 5 clone RP11-197113, WORKING DRAFT
SEQUENCE, 10 unordered pieces.
AC104112
AC104112.1 GI:17298613
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 157165)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 157165)
DOE Joint Genome Institute.
Direct Submission
Submitted (04-DEC-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Summary Statistics
Consensus quality: 146558 bases at least Q40
Consensus quality: 154071 bases at least Q30
Consensus quality: 154899 bases at least Q20
Estimated insert size: 169320; agarose-fp estimation
Estimated insert size: 156265; sum-of-contigs estimation
Quality coverage: 11.31 in Q20 bases; agarose-fp estimation
Quality coverage: 12.26 in Q20 bases; sum-of-contigs estimation
NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1541: contig of 1541 bp in length
* 1542 1641: gap of unknown length
* 1642 3077: contig of 1436 bp in length
* 3078 3177: gap of unknown length
* 3178 5571: contig of 2394 bp in length
* 5572 5671: gap of unknown length
* 5672 7989: contig of 2318 bp in length
* 7990 8089: gap of unknown length
* 8090 15847: contig of 7758 bp in length
* 15848 15948: gap of unknown length
* 15949 24938: contig of 8991 bp in length
* 24939 25038: gap of unknown length
* 25039 38970: contig of 13932 bp in length
* 38971 39070: gap of unknown length
* 39071 62670: contig of 23600 bp in length
* 62671 62770: gap of unknown length
* 62771 103332: contig of 40562 bp in length
* 103333 103432: gap of unknown length
* 103433 157165: contig of 53733 bp in length.
Location/Qualifiers
1. 157165
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-197113"
/clone_1db="RP11 human BAC library 11"

ORIGIN

Query Match 4.8%; Score 45.8; DB 2; Length 157165;
Best Local Similarity 52.3%; Pred. No. 3.4;
Matches 101; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 761 CATTAATCAACAAATCTACAGCGCTGAGAGTGAGAAACTGTGAGCAACACTTGGTC 820
DB 28132 CATCCAGAAACAAATCTTGGATCTTCAATCCAGTCAAGTAAACACTTAACCAATGAC 28191
QY 821 TTATTTAAATATCTAATCATGTAAATTTCAATTTATCTTCTGATCTTGCATCA 880
DB 28192 ACATTCAAAACCTACGTGCTTCTCTCTCTCAATCTTAATGTTTATGATGCA 28251
QY 881 CATCTCTTATTTCAAAATCTCACTTAAATCATATAATTAATGAGTTTATCAAA 940
DB 28252 TAACTCATATCTTATATAAATATATTATTATTAATGTTTATGATGTTATTTAAAT 28311
QY 941 TACTCATCTTGA 953
DB 28312 AATTTAACTTTTA 28324

RESULT 10

LOCUS AC116358 180797 bp DNA linear PRI 23-APR-2002
DEFINITION Homo sapiens chromosome 5 clone RP11-679B13, complete sequence.
AC116358
AC116358.2 GI:20270119
HTG.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 180797)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Submitted (27-MAR-2002) Production Sequencing Facility, DOE Joint

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

REFERENCE 3 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (23-APR-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Apr 23, 2002 this sequence version replaced gi:19745040.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40.99.9% of Sequence;
Estimated Total Number of Errors is 0.1.
NOTE: This insert is not the entire sequence of the clone (entire sequence is 206Kb). It is clipped at the overlap with AC114283. The number of bases overlapped is 113459.
location/Qualifiers
1. 180797
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-679B13"

ORIGIN
Query Match 4.8%; Score 45.8; DB 9; Length 180797;
Best Local Similarity 52.3%; Pred. No. 3.5;
Matches 101; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 761 CATATATACAAATATCTACAGCTGAGAGTCGAAACTGTCCAGCAGCACTTGTGTC 820
DB 11216 CATCGAAACATCTTGTGATCTCTTCATCCAGTCAAGTTAACATTAACCATATAC 11275
QY 821 TTAATTAATATCTTAACATGTAATTTCAATTTATCTCTTCACTTTGCTATCA 880
DB 11276 ACATTCAAACCTACTGCTGTGCTCTCTCTTCACATACCTTAATGTTTGTATGCA 11335
QY 881 CATGCTCTTTATTTCAAAATCTCACTTAATTCATTAATAATAGGTTTATCAAA 940
DB 11336 TAAAGCATATCTTTATTAATAATATTAATTTATTAATGTTATGTTATTTTAAAT 11395
QY 941 TACATCATCTTGA 953
DB 11396 AATTAACTTTA 11408

RESULT 11
AC051636/c 199421 bp DNA linear HTG 17-JUN-2000
LOCUS Homo sapiens chromosome 5 clone RP11-679B13 map 5, WORKING DRAFT
DEFINITION
SEQUENCE, 24 unordered pieces.
AC051636
AC051636.3 GI:8570502
VERSION HTG; HTGS PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 199421)
AUTHORS Birren, B., Linton, L., Nussbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 5, clone RP11-679B13
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 199421)
Birren, B., Linton, L., Nussbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F.,
Bogosavlavic, L., Bouhassira, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Chopel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., DeAngelis, K., Dewar, K., Diaz, J., S.,
Dodge, S., Domono, M., Doyle, M., Ferreira, P., Fitzhugh, N., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Labocque, K., Lamazares, R., Landers, T., Lehoczy, J.,

TITLE
JOURNAL
COMMENT
Levine, R., Liu, G., Locke, K., Macdonald, P., Margulis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPherson, R.,
Meldrum, J., Menus, L., Milova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neil, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Strange-Thomson, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Teefaye, S., Theodore, V., Tirrell, A., Travers, M., Trigilio, J.,
Vaasilev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (15-APR-2000) Whitehead Institute/MIT Center for Genome
Research 320 Charles Street, Cambridge, MA 02141, USA
On Jun 17, 2000 this sequence version replaced gi:8099804.
All repeats were identified using RepeatMasker:
Smith, A. F. A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6086
Center clone name: 679 B.13
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 188667 bases at least Q40
Consensus quality: 193870 bases at least Q30
Consensus quality: 195863 bases at least Q20
Insert size: 208000; agarose-fp
Insert size: 197121; sum-of-contigs
Quality coverage: 4.3 in Q20 bases; agarose-fp
Quality coverage: 4.5 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1243: contig of 1243 bp in length
* 1244 1343: gap of 100 bp
* 1344 2579: contig of 1236 bp in length
* 2580 2679: gap of 100 bp
* 2680 4199: contig of 1520 bp in length
* 4200 4299: gap of 100 bp
* 4300 5975: contig of 1676 bp in length
* 5976 6075: gap of 100 bp
* 6076 8547: contig of 2472 bp in length
* 8548 8647: gap of 100 bp
* 8648 11168: contig of 2521 bp in length
* 11169 11268: gap of 100 bp
* 11269 14811: contig of 3543 bp in length
* 14812 14911: gap of 100 bp
* 14912 18027: contig of 3116 bp in length
* 18028 18127: gap of 100 bp
* 18128 20472: contig of 2345 bp in length
* 20473 20572: gap of 100 bp
* 20573 25056: contig of 4484 bp in length
* 25057 25156: gap of 100 bp
* 25157 28977: contig of 3821 bp in length
* 28978 29077: gap of 100 bp
* 29078 33117: contig of 4040 bp in length
* 33118 33217: gap of 100 bp
* 33218 35890: contig of 2673 bp in length
* 35891 35990: gap of 100 bp
* 35991 42767: contig of 6777 bp in length

```
* 42768 42867: gap of 100 bp
* 42868 49422: contig of 6555 bp in length
* 49423 49522: gap of 100 bp
* 49523 57661: contig of 8139 bp in length
* 57662 57761: gap of 100 bp
* 57762 66485: contig of 8724 bp in length
* 66486 66585: gap of 100 bp
* 66586 79774: contig of 13189 bp in length
* 79775 79875: gap of 100 bp
* 79876 93302: contig of 13428 bp in length
* 93303 93402: gap of 100 bp
* 93403 108149: contig of 14746 bp in length
* 108149 108248: gap of 100 bp
* 108249 127542: contig of 19294 bp in length
* 127543 127642: gap of 100 bp
* 127643 148183: contig of 20541 bp in length
* 148184 148283: gap of 100 bp
* 148284 168177: contig of 19894 bp in length
* 168178 168277: gap of 100 bp
* 168278 199421: contig of 31144 bp in length.
```

FEATURES

```
Source
1..199421
  /organism="Homo sapiens"
  /mol_type="genomic DNA"
  /db_xref="taxon:9606"
  /chromosome="5"
  /map="5"
  /clone="RP11-679B13"
  /clone_id="RPC1-11 Human Male BAC"
1..1243
  /note="assembly_fragment"
  /note="assembly_fragment"
1344..2579
  /note="assembly_fragment"
2680..4199
  /note="assembly_fragment"
4300..5975
  /note="assembly_fragment"
clone_end:SP6
vector_side:right"
6076..8547
  /note="assembly_fragment"
8648..11168
  /note="assembly_fragment"
11269..14811
  /note="assembly_fragment"
14912..18027
  /note="assembly_fragment"
18128..20472
  /note="assembly_fragment"
20573..25056
  /note="assembly_fragment"
25157..28977
  /note="assembly_fragment"
29078..33117
  /note="assembly_fragment"
33218..335890
  /note="assembly_fragment"
clone_end:r7
vector_side:left"
35991..42767
  /note="assembly_fragment"
42868..49422
  /note="assembly_fragment"
49523..57661
  /note="assembly_fragment"
57762..66485
  /note="assembly_fragment"
66586..79774
  /note="assembly_fragment"
79875..93302
  /note="assembly_fragment"
93403..108148
  /note="assembly_fragment"
108149..127542
  /note="assembly_fragment"
127543..148183
  /note="assembly_fragment"
148184..168177
  /note="assembly_fragment"
168178..199421
  /note="assembly_fragment"
```

```
misc_feature 108249..127542
              /note="assembly_fragment"
misc_feature 127643..148183
              /note="assembly_fragment"
misc_feature 148284..168177
              /note="assembly_fragment"
misc_feature 168278..199421
              /note="assembly_fragment"
```

ORIGIN

```
Query Match 4.8%; Score 45.8; DB 2; Length 199421;
Best Local Similarity 52.3%; Pred.No. 3.5;
Matches 101; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
```

```
Gy 761 CATTAATCAACAAATCTACAGCTGGAGAGTGGAAAACGTGTGAGCAACACTTGTC 820
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 68387 CATCAGAAACAAATCTTGATCTTCATCCAGTCAAGTAAACATTAACCATGAC 68328
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Gy 821 TTATTTAAATATTTCAATGATTAATTTCAATTTACTTTCTGATCTGCTATCA 880
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 68327 ACATTCAAAACCTTACGCTGCTGCTCTCTTCTCAATCTTAATGTTTATGTC 68268
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Gy 881 CAGTCTCTTATTTCAAAATCTCATTAAATCAATTAATGAATGAGTTATCAAA 940
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 68267 TAAGTCATATCTTATATAAATATATTATTAATGTTTATAGTTATGTTATTAAT 68208
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Gy 941 TACATCATCTTGA 953
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 68207 AATTAACTTTTA 68195
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 12

AL672013

LOCUS AL672013 207905 bp DNA linear ROD 23-MAY-2002
DEFINITION Mouse DNA sequence from clone RP23-32G8 on chromosome 11, complete
sequence.

ACCESSION

AL672013

VERSION

AL672013.9 GI:21104143

KEYWORDS

HTG.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

REFERENCE

Oliver, K.

AUTHORS

Submitted (23-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,

TITLE

Cambridgehire, CB10 1SA, UK. E-mail enquiries:

JOURNAL

humquer@esanger.ac.uk Clone requests: clonerequests@sanger.ac.uk

COMMENT

On May 22, 2002 this sequence version replaced gi:20794189.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest. The following

abbreviations are used to associate primary accession numbers given

in the feature table with their source databases: Em., EMBL; Sw.,

SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP

database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-32G8 is

from the RP23-Mouse PAC Library

constructed by the group of Pieter de Jong.

For further details see http://www.choxi.org/bacpac/home.htm

VECTOR: PBAC3.6.

Location/Qualifiers

FEATURES

/protein_id="AA051558.1"
/db_xref="gi:28828978"
/translation="MARNQNRSGSTTDLRNLSSTSGCGNINISGCGNQDLAI
LSPKSSLSMSWSSSYNLNTLSSSKSPSLTLINQDNTTTTSTNGALISEQO
QLSMIPRYQKQYEDVKILFERNTEPMQMKRSEIDKMSQRLNIGKRYD
LVESDIAIVKMKSTELIDNLMOSELKOPSEKRNHFRAGQDNILKNGEKIDK
KISIFYSCKFLIDIPYIWRSLSDNDFYFESKLTIEKOWTSMKQPPKLTGYKRL
NESTRILGPIREKYSISLTILPEKSKETNEBELSRSLPNSILSKDQKRL
QOTIEKMFPLKMSIYITWLFYPRKQTNNNNNDPDEBQDEKILSDYQSKS
KFNKSLQKYTSSLSMSSTTTTNSPTFESKINIKPVSEKLNQLPLNFTSLSEST
FNMSPYLKECLKFKYQDVNSGELSINSQSGQGDYODINNNNNNSNGSDSQOI
EYSLIDFENSNFIPIKRTSEWIEIIEDFLANVQKFLIDIKSAKELSRSEIPF
ILDFKNIHLFNNHKOILYHQILQLIO"
join(13607..15915,16123..16126)
/note="GeneID exon scores (in order of location ranges):
184.62, -1.82 - GSCJ_ID dd_00635"
/codon_start=1
/product="similar to Plasmodium falciparum (isolate 3D7).
Hypothetical protein"
/protein_id="AA051559.1"
/db_xref="gi:28828979"
/translation="MFSILCGDKMYFLNIFEDIFLVKBSIIMDSFRINLSIQSD
IFSOIKSBDKMFSEFLWYNDDPIQSIKNTNGITPSNELFLKINQLYNNIKSDPI
YLLDNNINININNN
YKSLNEFTSTQDRIDOLILNKKTKTSTTSLNKKDELIFSLSKLSKIPYKHIVN
PNLYPLNSINIPDTTNSNCTSPSKNSINILDTISSPPPPSPHSSPSIQRHN
NNNNNNNNNNPTPTNNNNSEILKESLPTIDLKQPYGCTVWNEVGVYSOI
LKQDLFNHNDWDRIKTEWKHIIQISNGHNHNDANDGGDDAVWNEHSTII
YIPQSPPTITSYMSISLEISKFLNIDNLIPIESITLWLFPIVNDLSTST
STTASNTLSTNNNTTITIKENEGYIQLIDLKIGFLFSELENSSSSKPTSPS
NVIIKDSIFPKAKSHOISIISTKNVDENIELOKQOQOQOQOQOQOQOQOQOQOQ
QOQ
FVYHAKSVKPKQ
TNKFQSLFSDITQITDALATAPTTIPSTNSISQSSANNSALASPISTIDYRNQOQ
QOQNNNQOQSLSPSLSGHMQYFIHG"
join(16453..17015,17101..17404,17491..17972,18042..18717)
/note="GeneID exon scores (in order of location ranges):
15.09, 20.07, 28.00, 42.31 - GSCJ_ID dd_00636"
/codon_start=1
/product="Hypothetical protein"
/protein_id="AA051560.1"
/db_xref="gi:28828980"
/translation="MPKPSYSLLAHLKOCVEKSKDHTTAIVDYINCDSLEIEB
SODKINNOSIKKEIPIYSLNLTAKCTQVLTANNASCTPTASNDTPPTPNP
TPAPESILKKKRPILPSTNGVINSCEKTKTYIDPRECEBKMKMLIGQITLY
GTRSGGTTTSTIVCELSIKGHSIFIDQGVINSIDPMFVLIINSIMQSLIDT
PTTKEIDEMSIROLKDLITFNRRKSLFNKQVHFLIDEFNIDGGEETTOVLSDF
REMNGOSSLAIKSLAVIGTFSIDPNTSNTSPENIAOSFKVNLFOKEVELFLFK
FOKSGVEIYPOVDIDFITNGHKGIVNVEGIIOTSLSEFNITKNWMDHILNKOLAT
KLKRYTVMTRCLNTGSDIPFKTIDILFKSYSTPTPSKTESYLAHEGILSD
IEEINVPSSNIIKNIVLNHLRKTAVIGSVPTGDKKEDIVQQLINAAKQISPL
IILIAIORSFKCTIRTOFOQKVCCEIYNESIVLTLISRRANSNSYISPN
NIIYNKTYRVDIVIGNELGRVYLEFVALTSMEDVNGIERLVIENYCATGECM
VIHFTWVRSYDYOYPRVYKTVGIIHIDHKYSNLEETIYPTQOS"
complement(join(19030..19693,19776..20221,20305..20619,
20747..21433))
/note="GeneID exon scores (in order of location ranges):
69.01, 37.05, 31.88, 80.39 - GSCJ_ID dd_00629"
/codon_start=1
/product="Hypothetical protein"
/protein_id="AA051561.1"
/db_xref="gi:28828981"
/translation="MSKFTNLRKSSNMNTGTAGRSSVYIDTSGNNGSGSPDT
SIINTPGEALNEKVIPESEIKLSKNNLIDIKSKNRYRGGEFATQVK
MSOSFSDYKSTTHSSAQOYVQOQOQAAANNGSTSDNLGFLALGPTSS1
TKTSGFSEWINSWEGIKPIOLIGIDKQYRKKPDQAVOENYISIKIGOOT
OKTIDILKIYVKEKSKLKNVENVKQYVYVLTPEENWHTFELDLIYAVESML
LNGAVAYAGIKTYIISLRTWCANBQDFPKSETERDARVLELQCEBDIKHOPID
LVLPFYFLKLSIEIRKNEFLPPPPPPNTYIVPPPIHLPVLVAFERRQSLSEL
LIMVQSDLPNISMGTFLPYDIVSCETIEIANLPSTPLAKTLPOSLINININHD
NYRINTQOQITNLIAFEFYVNIIFQSGCEYLPPEPKASIIQNGNLISGSGSP1
GCLFSRVAPVARPHOHLFVYIPSOALEVALPSTWNTVNGTNSPASTILGSG
QANESMAKRVITTFPDQVQSDLTENDSSITVLNAYQODIPIYQSFIRKHYISG

CDS
SYLSRDEQSLVFTHALGSLPADPPEPKQSNNNNSKXSSSSSTPTQHPSPALS
PHSLONTNANSQDENIED"
join(23269..24061,24158..29665,30674..30722,32178..32246,
32344..33160,33273..33423)
/note="GeneID exon scores (in order of location ranges):
79.45, 607.64, -1.34, 4.63, 78.55, 12.61 - GSCJ_ID
dd_00630"
/codon_start=1
/product="similar to Plasmodium falciparum (isolate 3D7).
Query Match 4.7%; Score 45.2; DB 3; Length 38692;
Best Local Similarity 50.5%; Pred. No. 4;
Matches 110; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
Oy 735 GACTGCTCCCAAGCAAGAGCATGATTAATACAAAATCTCACAGCTGAGAGTC 794
Db 14102 GAATTTCAACATCACAACAAGATGATCAATTAATTTGAATTAACAAATAAA 14161
Oy 795 GAAGACTGCGAGCAGACACTGGCTTATTTAAATATTTCAACATGAAATTTCAATTTA 854
Db 14162 ACAACTGTGTACACCACTAGCAGCTAATTTAAGAGATGAAATTTATTCATTTCA 14221
Oy 855 TCTCTTACTTTCGATCTGCTATCATCATGCTCTTATTTCAAAATCTCACATTAAT 914
Db 14222 AAACCTTCAAAAATATTTTAATAACATATGTTAATATCAAACTTTATTTTAAAT 14281
Oy 915 TCATATTAATTAATGGTTTATTCAAATACATCATCTTG 952
Db 14282 TCATATTAATTAATTCGATACACAAATATTTCAAAATG 14319
RESULT 14
AC087834
LOCUS AC087834 165062 bp DNA linear PRI 22-MAY-2004
DEFINITION Pan troglodytes clone RP43-124117, complete sequence.
AC087834
AC087834.2 GI:15144422
VERSION
KEYWORDS
SOURCE Pan troglodytes (chimpanzee)
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL
COMMENT
Submitted (09-ANG-2001) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 165062)
Green, E.D.
Direct Submission
Submitted (09-ANG-2001) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Gaithersburg, MD 20877, USA
4 (bases 1 to 165062)
Green, E.D.
Direct Submission
Submitted (12-MAY-2004) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Gaithersburg, MD 20877, USA
On Aug 9, 2001 this sequence version replaced gi:12597741.
----- Genome Center
Center: NIH Intramural Sequencing Center
Web site: http://www.nisc.nih.gov

Contact: nisc.zoo@nhgri.nih.gov
 ----- Project Information
 Center project name: and
 Center clone name: 124117

This sequence was finished as follows unless otherwise noted:
 all regions were double-stranded, sequenced with an
 alternate chemistry, or covered by high quality data
 (i.e., phred quality >= 30); an attempt was made to resolve
 all sequencing problems, such as compressions and repeats;
 all regions were covered by at least one plasmid subclone
 or more than one M13 subclone; and the assembly was confirmed
 by restriction digest.

CLONE LENGTH: This sequence represents the entire insert of
 this clone unless otherwise noted. If there are overlapping
 clones, the overlaps are noted in the beginning and end of
 the Features section.

FEATURES

source

Location/Qualifiers

1..165062
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 /clone="RP43-124117"
 /clone_id="RP43"

misc_feature

/note="BAC resource: <http://bacpac.choiri.org/>"
 <1..52782

misc_feature

/note="This sequence is not the entire insert of clone
 RP43-124117, clone overlaps with GenBank Accession Number
 AC087730 (nucleotides 73132-148293) clone RP43-89H17
 (center project name anc); this annotated segment
 represents overlap with nucleotides 95501-148293 of
 AC087730"
 109343..109405
 /note="single clone coverage"

ORIGIN

Query Match 4.7%; Score 45.2; DB 9; Length 165062;
 Best Local Similarity 52.7%; Pred. No. 4.8;
 Matches 98; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

759 AGCATTAATCAACAATAATCTACCGCTGAGAGTCGAAAACCTGTCGACGACCACTTGG 818
 |||||
 108034 AGTAATAATTAATAAATAGCAAGCACTAAAGAAAATACCAAAACCAAGCATTTTG 108093
 |||||
 819 TCTTAATTAATAATCTTAACATGTAATTTCAATTTATCTCTTATCTTGTGATCTGAT 878
 |||||
 108094 AATTAGTCAATGAATCTAAAGAAAGATGATTTTTTTTTTGAGATTCGTCTGAGTGT 108153
 |||||
 879 CACATGCTCTTAATTTCAAAAATCTCACTTTAAATTCATATAATPAATGGTTTATCA 938
 |||||
 108154 GGTAATATGTAATTTTGGCAAAACTATTAACAAATTTGGCAAAATGAAAGAAATATTTTA 108213
 |||||
 939 AATACA 944
 |||||
 108214 AAAACA 108219

RESULT 15
 AC146421/c 178824 bp DNA linear PRI 19-MAY-2004
 LOCUS AC146421
 DEFINITION Pan troglodytes BAC clone RP43-27017 from 7, complete sequence.
 ACCESSION AC146421
 VERSION AC146421.2 GI:46395435
 KEYWORDS HTG.
 SOURCE Pan troglodytes (chimpanzee)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
 REFERENCE 1 (bases 1 to 178824)
 AUTHORS Nguyen,C., Haglund,K. and Meyer,R.
 TITLE The sequence of Pan troglodytes BAC clone RP43-27017
 JOURNAL Unpublished (2001)

REFERENCE

2 (bases 1 to 178824)
 WILSON,R.K.
 Direct Submission
 Submitted (13-AUG-2003) Genetics, Genome Sequencing Center, 4444
 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE

3 (bases 1 to 178824)
 WILSON,R.K.
 Direct Submission

REFERENCE

Submitted (16-APR-2004) Genetics, Genome Sequencing Center, 4444
 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE

4 (bases 1 to 178824)
 WILSON,R.K.
 Direct Submission

REFERENCE

Submitted (19-MAY-2004) Washington University School of Medicine,
 Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO
 63108, USA

COMMENT

On Apr 16, 2004 this sequence version replaced gi:33620996.
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu>
 Contact: submissions@wustl.edu
 ----- Summary Statistics
 Center project name: C_PT027017

NOTICE:

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by sequence
 from more than one subclone; and the assembly was confirmed by
 restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren,
 Department of Genetics, Washington University, St. Louis MO. For
 additional information about the map position of this sequence, see
<http://genome.wustl.edu>

SOURCE INFORMATION:

The RCI-43 BAC library has been constructed by Chung-Li Shu. DNA
 was isolated from white blood cells obtained from a male chimpanzee
 (Pan troglodytes, 'Clint', Yerkes #C0471; birthdate: 6-6-80). The
 clone and detailed information can be obtained from Resgen
 (<http://www.resgen.com>) or Pieter de Jong and co-workers at
<http://www.bacpac.choiri.org>.

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

FEATURES

source

Location/Qualifiers
 1..178824
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 /chromosome="7"
 /map="7"
 /clone="RP43-27017"
 /clone_id="RCI-43"
 25545..25790
 /note="Sequence derived from one plasmid subclone."

ORIGIN

unsure

Query Match 4.7%; Score 45.2; DB 9; Length 178824;
 Best Local Similarity 52.7%; Pred. No. 4.9;
 Matches 98; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

759 AGCATTAATCAACAATAATCTACCGCTGAGAGTCGAAAACCTGTCGACGACCACTTGG 818
 |||||
 160054 AGTAATAATTAATAAATAGCAAGCACTAAAGAAAATACCAAAACCAAGCATTTTG 159995

Qy	81.9	TCTTAATTTAAATAATCTACAGTGAATTCCATTATCTCTACTTCGTGATCTTGCTAAT	878
Dd	159994	AATTAGTCAATTGAATCAAAAAGAAGTAGATTTTTTTTTGAGATTCTGGTCTAGGTGT	159935
Qy	87.9	CACATGCTCCTAATTCCAAAAATCCACTTTAAATTCATATATAAATATGGGTTATTC	938
Dd	159934	GGTATATGTGTATTTTTCGAAAAACTATTAACAATTGTGGCAAATGAAGGAAATATTTTA	159875
Qy	93.9	AATACA	944
Dd	159874	AAAACA	159869

Search completed: April 9, 2005, 05:16:00
Job time : 4378 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_p2n model

Run on: April 9, 2005, 08:00:52 ; Search time 3511 Seconds
(without alignments)
2981.394 Million cell updates/sec

Title: US-09-087-136-1
Perfect score: 1462
Sequence: 1 MSEIDPLAEFLPEDGDGRNA.....INKIYQGESKTVQALGLI 275

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=exh
-O=/cgn2_1/USPTO.spool/US09087136/runat_07042005_163616_919/app_query.fasta_1.455
-DB=EST -QFMT=fastap -SUFFIX=rc -MINMATCH=0.1 -LOOPCL=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blowsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pic -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09087136.@CGN 1.1.4352.@runat_07042005_163616_919 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_g981:*
9: gb_g982:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	1113	76.1	752	1	AU218734
2	1002	68.5	666	6	CB102172 MMV_SQ005
3	991	67.8	559	3	CB386085 OSTF038D2
4	933	63.8	614	6	CB102170 MMV_SQ005
5	888	60.7	605	6	CB102173 MMV_SQ005
6	872	59.6	584	6	CB102177 MMV_SQ005
7	860	58.8	586	6	CB102175 MMV_SQ005
8	847	57.9	562	6	CB102169 MMV_SQ005
9	718	49.1	495	6	CB102055 MMV_SQ001

10	689	47.1	485	6	CB102059
11	639	43.7	377	1	AV202470
12	619	42.3	360	7	D74918
13	619	42.3	438	6	CB102066
14	615	42.1	433	6	CB102067
15	605	41.4	425	6	CB102064 MMV_SQ001
16	605	41.4	425	6	CB102065
17	600	41.0	360	7	D75085
18	585	40.0	360	7	AV203242
19	585	40.0	360	7	D69562
20	526	36.0	366	6	CB102062
C 21	444	30.4	300	6	C55654
C 22	439	30.0	300	1	AV200339
C 23	436	29.8	300	1	AV199733
C 24	350.5	24.0	653	6	CA958252
C 25	346	23.7	339	7	D65944
26	206.5	14.1	633	2	BE13546
27	188	12.9	728	1	A1124451
28	183.5	12.6	476	4	BM774830
29	137.5	9.4	757	3	AY432342
30	119	8.1	746	8	A2977868
31	116	7.9	722	5	BX623602
C 32	114	7.8	729	7	CR527015
33	114	7.8	773	7	CR527016
34	111.5	7.6	1197	3	CNS0A9LR
35	111.5	7.6	1703	4	BM558476
C 36	110.5	7.6	819	7	CO163741
37	110.5	7.6	1015	5	BQ712448
38	110.5	7.6	1310	5	BUI72489
39	109.5	7.5	997	4	BG964389
40	109	7.5	612	7	CV164818
41	109	7.5	1253	3	CNS0ADNL
C 42	108	7.4	782	7	CO163663
C 43	107	7.3	180	7	D72252
C 44	107	7.3	1438	8	AQ012204
45	106	7.3	870	7	CO922299

ALIGNMENTS

RESULT 1
LOCUS AU218734/C 752 bp mRNA linear EST 17-JUL-2001
DEFINITION AU218734 unpublished oligo-capped cDNA library, stage 11
ACCESSION AU218734
VERSION AU218734.1 GI:14856891
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 752)
Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
A complementary view of the C.elegans genome
Unpublished (2001)
CONTACT: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers
1. .752
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="YK86801"
/sex="hermaphrodite"
/tissue_type="whole animal"

/dev stage="L1"
/clone_lib="unpublished oligo-capped cDNA library, stage
L1"

ALIGNMENT SCORES:

Pred. No.: 3,116-100 Length: 752
Score: 1113.00 Matches: 210
Percent Similarity: 99.53% Conservative: 0
Best Local Similarity: 99.53% Mismatches: 0
Query Match: 76.13% Indels: 1
DB: 1 Gaps: 0

US-09-087-136-1 (1-275) x AU218734 (1-752)

```
QY 66 AsnleuLeuThrhr-1lelyshishsProsergiullelelyvalleuProgiuNs 85
DB 751 AATCTGCTGCAACCCATAAACAATCATCCATCGAATAATTTGAGTACTCCACAGGA 692
QY 85 PLYrThrArgAlaSPglugluProgiYArGInGIYArGProProGIYArGProArGIY 105
DB 691 TTAATCTCGTGTGATGAGAGAGCCCGGCCAGAGAGCTCCACGATCGCCCTGTGA 632
QY 105 sMetProArGHIsgIuSerSerThiSerleuWetGIuSerProArGIYrThMetThAr 125
DB 631 GATGCCGCGTCAGAACTTCACTTACATGATCACCAGCAGCATATGACTG 572
QY 125 GAAPSerIYsIleMePheGluLeuArGIYsPProPheGIuMetIleAlGIYArGPh 145
DB 571 TGATTTTAAATTAATGTTGAATGCGTGAACCAATTCGAATGATAGCTGACGTTT 512
QY 145 eGIuGIuGIuYrSerleuGIYArGAlaTrpVallySGIYHISMeCaenSngIuTYrGI 165
DB 511 TGAAGAGAATATTCTACTTGTAAGCATGGGTAAAGACACATGAATATGATATGA 452
QY 165 uProIleYsAlaGInArGThrAsPTrAlaProAsnleuAlaValAsPTrYrleuAlaCY 185
DB 451 ACCAATAAAGCTCAAGAGACAGACTATGACCGAATCGCTGTGTGATTAATCTTGATG 392
QY 185 sArGIuIleHISArGMeProArGProAsPTrYrSerIlePProGIuLeuProIleValPr 205
DB 391 TGGCGAGATTCATCGAATGCCACGTCACATTAATCAATTCCTGAGCGCAATTTGTC 332
QY 205 oSerArGIleAsPgluPheAsPAlaThrValAsPProArGIYrGIuThrAsPleuYsAs 225
DB 331 ACTTGAATTCAGTGAATTCAGCGTACAGTCATCAAGATATGAAAACAGATTTGAAAA 272
QY 225 nGIuTYrIleArGHIstRPySGInVallySGIYTrPySAlaHISGInArGAr 245
DB 271 TGAATACATTCGTCATGTGAACAAGTCAAAAAGGTTGGTGTCTCATCAACGTCG 212
QY 245 gThrAlaProHISAlaArGSerIleAlaIleuIleAsnlySIIeTYrGIuProGIYGIuSe 265
DB 211 GACTCTCCCTCCATGCAAGAGATGATGATATCAACAAATCTTACAGAGCTGAGAGTC 152
QY 265 rlyeThrValGIuGInAlaLeuGIYleuIle 275
DB 151 GAAAACTGTCCAGCAAGCACTTGCTTATT 121
```

RESULT 2
CB102172 666 bp mRNA linear EST 28-JAN-2003
LOCUS CB102172
DEFINITION MW S0005086 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA sequence.
ACCESSION CB102172
VERSION CB102172.1 GI:27927979
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryote; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.
REFERENCE Walhout,A.J., Sordella,R., Lu,X., Hartley,J.L., Temple,G.F.,
AUTHORS Brasch,M.A., Thierry-Mieg,N. and Vidal,M.

TITLE Protein interaction mapping in C. elegans using proteins involved
in vulval development
JOURNAL Science 287 (5450), 116-122 (2000)
MEDLINE 20082953
PUBMED 10615043
COMMENT Contact: Vidal M
Marc Vidal Laboratory
Dana Farber Cancer Institute
1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
Tel: 617 632 5180
Fax: 617 632 5739
Email: Marc.Vidal@fci.harvard.edu
Trace dnp10318.x with Bait lin-53
POLYA=No.

FEATURES

source

Location/Qualifiers

1..666
/organism="Caenorhabditis elegans"
/mol_type="mRNA"

/strain="N2"

/db_xref="taxon:6239"

/sex="Hermaphrodite and male"

/tissue_type="whole animal"

/dev stage="mixed stage"

/clone_lib="AD-wrmcDNA"

/note="The AD-wrmcDNA library was generated with poly(A)+
RNA isolated from both hermaphrodite and male N2 worms of
all larval stages, embryos, adults and dauers and the
subsequent generation of cDNAs by poly(A) priming. The
cDNAs were cloned into pC86"

ORIGIN

ALIGNMENT SCORES:

Pred. No.: 2,896-89 Length: 666
Score: 1002.00 Matches: 188
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 98.95% Mismatches: 0
Query Match: 68.54% Indels: 0
DB: 6 Gaps: 0

US-09-087-136-1 (1-275) x CB102172 (1-666)

```
QY 1 MetSerGIuIleAsPProleuAlaGIuPheLeuProGIuAsPglYAsArGAsnAla 20
DB 95 ATGTGAGAAATGATCATCTTGCAGATCTTCTTCCGAAAGACGAGATGAAATGCT 154
QY 21 ArgGlnAsnAsPProleuIleSerGIYGIYProleuProleuGIuSerProSerArGIYs 40
DB 155 CGTCAAAATGATCCATGTGATTAAGCGAGGTCACATTCATGGAATCGCACAAGAAAA 214
QY 41 leuThrSerleuIleuSerTYrAsPProThrValProGIuSerProAsMetIlyPheAla 60
DB 215 CTCACATCCCTGTATCTTATATCCAGACATTCGAGATCCGGAATATGAAATTCGCC 274
QY 61 ArgLYsArGleuGIYAsnleuLeuThrThrIleYsHISAsPProSerGIuIleleGIY 80
DB 275 AGHAAAAGCTGTGGAAATCTGCTGACACCATTAACATCACTCGGAATATTTGGA 334
QY 81 ValIleuProGIuAsPTrThrTrArgAlaAsPgluGIuPProGIYArGInGIYArGProPro 100
DB 335 GTACTCTCCAGAAATTAATCTGTCTGATGAAGAGCCCGGCCAGAGAGCTCCACACA 394
QY 101 GIYArGProArGIYMePProArGHIsgIuSerSerThiSerleuWetGIuSePProArGI 120
DB 395 GGTGCCCGTCGTAAATATGCGCGTCAACAAATCTTCACTTATGGAATCACACGCG 454
QY 121 LysThrMetThrArGAsPserIYsIleMePheGIuLeuArGIYrYsPProPheGIuMet 140
DB 455 AGGACTATGACTCGATTTAAATTAATTTGTTGAATGCTGGAAAAACATTCGAAATG 514
QY 141 IleAlaGIYArGProPheGIuGIuGIuTYrSerleuGIYArGAlaTrpVallySGIYHISMet 160
DB 515 ATAGCTTGACGTTTGAAGAAAGAAATTAATTCATCTGTGTAGAGATGAGTTAAAGACACATG 574
```

Qy 161 AaahngluTygluProleuLysalaIngrThraPryAlaProAsnleu1a1a1 180
Db 575 AATATGATATGACCAAGTCTAAAGCTCAAGAGACAGACTGACCAATCTGGCTGT 634
Qy 181 AaPryleuAlaCyAarglu1leH1aarg 190
Db 635 GATTATCTTCATGTCGCGAGATTCATCGA 664

RESULT 3
CB386085 559 bp mRNA linear EST 15-MAY-2003
LOCUS OSF038D2_1 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA sequence.
DEFINITION CB386085
ACCESSION CB386085.1 GI:30727795
VERSION EST.
KEYWORDS Caenorhabditis elegans
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida;
Rhabdilitida; Rhabdilitidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 559)
AUTHORS Reboul,J., Vaglio,P., Rual,J.F., Lamesch,P., Martinez,M.,
Armstrong,C.M., Li,S., Jacotot,L., Bertin,N., Janky,R., Moore,T.,
Hudson,J.R., Hartley,J.L., Brasch,M.A., Vandenhaute,J., Boulton,S.,
Endress,G.A., Jenna,S., Chever,E., Papasotiriopoulos,V.,
Tolias,P.P., Placek,J., Snyder,M., Huang,R., Chance,M.R., Lee,H.,
Doucette-Stamm,L., Hill,D.E. and Vidal,M.,
C. elegans ORFome version 1.1: experimental verification of the
genome annotation and resource for proteome-scale protein
expression
Nat. Genet. (2003) In press
JOURNAL Contact: Vidal M
COMMENT Marc Vidal Laboratory
Dana Farber Cancer Institute
1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
Tel: 617 632 5180
Fax: 617 632 5739
Email: Marc.Vidal@dfci.harvard.edu
Sequence tag of Gateway entry clones. The primers used were
designed on the predicted protein encoding ORF. C. elegans ORFome
cloning project : Contact david_hill@dfci.harvard.edu or
marc_vidal@dfci.harvard.edu
POLYA=No.

FEATURES
source Location/Qualifiers
1..559
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/sex="Hermaphrodite and male"
/tissue_type="whole animal"
/dev_stage="mixed stage"
/clone_lib="AD-wrmcDNA"
/note="The AD-wrmcDNA library was generated with poly(A)+
RNA isolated from both hermaphrodite and male N2 worms of
all larval stages, embryos, adults and dauers and the
subsequent generation of cDNAs by poly(A) priming. The
cDNAs were cloned into pPC86"

ORIGIN
Alignment Scores:
Pred. No.: 2,856-88 Length: 559
Score: 991.00 Matches: 186
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 67.78% Indels: 0
DB: Gaps: 0
US-09-087-136-1 (1-275) x CB386085 (1-559)

Qy 9 GluPheuleuProgluAaPglYAaPargAna1aarglInaAspProleu1leser 28
Db 1 GAGTTCTTGCTTCCAGAAAGACGAGATCGAATGCTCTCAAAATGATCCATTGATAGC 60

Qy 29 GlyGlyProleuProleuGluserProserArglyseuThraSerleuSerTyasp 48
Db 61 GGAAGTCCACTTCCATTGGATGCCAAGACGAGAAACATCATCTCTTATCTATGAT 120
Qy 49 ProThraValProgluSerProaspMetIysPhe1aarglyAargleu1Asnleu 68
Db 121 CCGACAGTTCGCGAGTCCAGATATGAAATTCGCCAGAAACGCTGGAAATCTGCTG 180
Qy 69 ThrThraIleuYsh1eH1aProserGlu1le1eGlyValleuProgluAaPryTyThra 88
Db 181 ACAACCACTTAAACATCCACCCATCGAAATATTTGAGTACTCCAGAAATTAATCTCTG 240
Qy 89 AlaaspGluGlupProgluAargGlu1AargProgluAargProgluAarglyMetProarg 108
Db 241 GCTGATGAAGAGCCCGGCGCGCAAGACGTCCACAGTCCCTCTGTAAGATCCGCT 300
Qy 109 HisGluSerSerThraSerleuMetGluSerProarglyThraMetThraArgAspSerIys 128
Db 301 CACGAATCTTCAACTTCATGGAATGCAACGCAAGACTATGACTCGATTCCTTAA 360
Qy 129 IleMetPheGluLeuAargGlyLysProPheGluMetIle1aGlyAargPheGluGlu 148
Db 361 ATTATGTTTGAATTCGCGAAACCAATTCGAATGATGCTGACGTTTGAAGAAAGAA 420
Qy 149 TySerleuGlyAargAlaTropVallyGlyHisMetAsnaengluTyrgluPro1le 168
Db 421 TATTCACCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
Qy 169 AlaGluAargThraPryTyra1ProAsnleu1aValaPryTyra1leu1aCyAargGlu1le 188
Db 481 GCTCAAGACAGACACTATGACCAAGATCTGCTGTGATATCTTCAATGCTGCGAGATT 540
Qy 189 HisArgMetProArgPro 194
Db 541 CATCGAATGCCACGCTCA 558

RESULT 4
LOCUS CB102170 614 bp mRNA linear EST 28-JAN-2003
DEFINITION MW_SQ005@B4 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA sequence.
ACCESSION CB102170
VERSION CB102170.1 GI:27927977
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida;
Rhabdilitida; Rhabdilitidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 614)
AUTHORS Walhout,A.J., Scordella,R., Lu,X., Hartley,J.L., Temple,G.F.,
Brasch,M.A., Thierry-Mieg,N. and Vidal,M.
Protein interaction mapping in C. elegans using proteins involved
in vulval development
Science 287 (5450), 116-122 (2000)
JOURNAL MEDLINE
PUBMED 20082953
10615043
Contact: Vidal M
Marc Vidal Laboratory
Dana Farber Cancer Institute
1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
Tel: 617 632 5180
Fax: 617 632 5739
Email: Marc.Vidal@dfci.harvard.edu
Trace dvp10316.x with Bait 1in-53
POLYA=No.

FEATURES
source Location/Qualifiers
1..614
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/sex="Hermaphrodite and male"
/tissue_type="whole animal"
/dev_stage="mixed stage"

/clone.lib="AD-wrmcDNA"
 /note="The AD-wrmcDNA library was generated with poly(A)+
 RNA isolated from both hermaphrodite and male N2 worms of
 all larval stages, embryos, adults and dauers and the
 subsequent generation of cDNAs by poly(A) priming. The
 cDNAs were cloned into pPC86"

ORIGIN

Alignment Scores:

Pred. No.:	1,896-82	Length:	614
Score:	933.00	Matches:	176
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	99.44%	Mismatches:	0
Query Match:	63.82%	Indels:	0
DB:	6	Gaps:	0

US-09-087-136-1 (1-275) x CB102170 (1-614)

```

QY      2 SerGluIleAspProLeuAlaGluPheLeuLeuProGluAspGlyAspArgAnaIArg 21
DB      82 GCAGAAATAGATCCACTGCGAGTCTTGCTCCAGAACGAGATCGAATGCTCGT 141
QY      22 GluAsnAspProLeuIleSerGlyGlyProLeuProLeuGluSerProSerArgLysLeu 41
DB      142 CAATATGATCATGATGATGAGCGAGTCCACTTCATTGGAATGCGCAAGCAAAAATCTC 201
QY      42 ThrSerLeuSerIleAspProThrValProGluSerProAspMetLysPheAlaArg 61
DB      202 ACATCCCTGTTATCTGATTCGACAGATCCGAGTCCGAGATGAGAAATTCGCGCAAA 261
QY      62 LysArgLeuGlyAsnLeuLeuThrThrIleLysHisHisProSerGluIleIleGlyVal 81
DB      262 AAACGCTCGGAAATCTGCTGACACCACTAAACATCACCACCATCGAATAATTTGGAGTA 321
QY      82 LeuProGluAspPylThrThrArgAlaAspGluGluProGlyArgGlnGlyArgProProGly 101
DB      322 CTCACGAGATATATATCTGCTGATGAGAGCCGCGCGCAAGAGATGACACAGGT 381
QY      102 ArgProArgLysMetProArgHisGluSerSerThrSerLeuMetGluSerProArgLys 121
DB      382 CGCCCTCGTAAAGATGCGCGTCAAGATCTTCACTTATGGAATCACCACGCAAG 441
QY      122 ThrMetThrArgAspSerLysIleMetPheGluLeuArgGlyLysProPheGluMetIle 141
DB      442 ACTATACCTCGATCTTAAATATGTTGAAATGCGGAAACCAATTCGAATGATA 501
QY      142 AlaGlyArgPheGluGluGluIleLysSerLeuGlyArgAlaIleProValLysGlyHisMetAsn 161
DB      502 GCTGACGCTTTGAGAGAAATATTCATTGCTGTAAGCATGGGTTAAAGGACACATGAT 561
QY      162 AsnGluIleGluProIleLysAlaGlnArgThrAspIleProAsnLeu 178
DB      562 AATGAAATATGAAACAATTAAGCTCAAGGACAGACTATGACCGAATCTG 612

```

RESULT 5
 CB102173 605 bp mRNA linear EST 28-JAN-2003

LOCUS MTU.S0005088 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA sequence.
 DEFINITION CB102173
 ACCESSION CB102173.1 GI:27927980

VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditioidea; Rhabditidae; Peloderiinae; Caenorhabditis.

REFERENCE
 1 (bases 1 to 605)
 Walhout,A.J., Sordeja,R., Lu,X., Hartley,J.T., Temple,G.F.,
 Brusch,M.A., Thierry-Mieg,N. and Vidal M.
 Protein interaction mapping in C. elegans using proteins involved
 in vulval development

TITLE
 JOURNAL
 MEDLINE
 PUBMED

20082953
 Science 287 (5450), 116-122 (2000)
 10615043

COMMENT

Contact: Vidal M
 Marc Vidal Laboratory
 Dana Farber Cancer Institute
 1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
 Tel: 617 632 5180
 Fax: 617 632 5739
 Email: Marc.Vidal@dfci.harvard.edu
 Trace dvp10320.x with Bait 11n-53
 POLYA=No

FEATURES

source

location/Qualifiers
 1..605
 /organism="Caenorhabditis elegans"
 /mol_type="mRNA"
 /strain="N2"
 /db_xref="taxon:6239"
 /sex="Hermaphrodite and male"
 /tissue_type="whole animal"
 /dev_stage="mixed stage"
 /clone_lib="AD-wrmcDNA"
 /note="The AD-wrmcDNA library was generated with poly(A)+
 RNA isolated from both hermaphrodite and male N2 worms of
 all larval stages, embryos, adults and dauers and the
 subsequent generation of cDNAs by poly(A) priming. The
 cDNAs were cloned into pPC86"

ORIGIN

Alignment Scores:

Pred. No.:	5,546-78	Length:	605
Score:	888.00	Matches:	168
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	60.74%	Indels:	0
DB:	6	Gaps:	0

US-09-087-136-1 (1-275) x CB102173 (1-605)

```

QY      1 MetSerGluIleAspProLeuAlaGluPheLeuLeuProGluAspGlyAspArgAnaIa 20
DB      100 ATGTCGAAATAGATCCACTGCGAGTCTTGCTCCAGAAAGCGAGATGGAATGCT 159
QY      21 ArgGluAsnAspProLeuIleSerGlyGlyProLeuProLeuGluSerProSerArgLys 40
DB      160 CGTCAAAATGATCATGTAATGAGCGAGTCCACTTCATTGGAATTCGCAAGCAAAA 219
QY      41 LeuThrSerLeuSerIleAspProThrValProGluSerProAspMetLysPheAla 60
DB      220 CTCACATCCCTGTTATCTGATTCGACAGTCCGAGTCAACGATATGAAATTCGCC 279
QY      61 ArgLysArgLeuGlyAsnLeuLeuThrThrIleLysHisHisProSerGluIleIleGly 80
DB      280 AGAAACGCTCGGAAATCTGCTGACACCACTAAACATCACCACATGGAATTAATGGA 339
QY      81 ValLeuProGluAspPylThrThrArgAlaAspGluGluProGlyArgGlnGlyArgProPro 100
DB      340 GTACTCCGAGAAATTAATCTGCTGATGAGAGCCGCGCGCAAGACGCTCACCA 399
QY      101 GLyArgProArgLysMetProArgHisGluSerSerThrSerLeuMetGluSerProArg 120
DB      400 GGTGACCTCGTAAAGATGCGCGTCAAGATCTTCACTTATGGAATCACCACGCG 459
QY      121 LysThrMetThrArgAspSerLysIleMetPheGluLeuArgGlyLysProPheGluMet 140
DB      460 AAGACTATGAGCTCGGAAATCTGCTGAAATATGTTGAAATGCTGGAACCAATTCGAATG 519
QY      141 IleAlaGlyArgPheGluGluGluIleLysSerLeuGlyArgAlaIleProValLysGlyHisMet 160
DB      520 ATAGCTGACGCTTTGAGAGAAATATTCATTGCTGTAAGCATGGGTTAAAGGACACATG 579
QY      161 AsnGluIleGluProIleLys 168
DB      580 AATATGATATGAAACAATTAAG 603

```

RESULT 6

CB102177
LOCUS CB102177 584 bp mRNA linear EST 28-JAN-2003
DEFINITION MW_S0005B12 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA
ACCESSION CB102177
VERSION CB102177.1 GI:27927984
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
AUTHORS Walhout,A.J., Scordella,R., Lu,X., Hartley,J.L., Temple,G.F.,
Brasch,M.A., Thierry-Mieg,N. and Vidal,M.
TITLE Protein interaction mapping in C. elegans using proteins involved
in vulval development
JOURNAL Science 287 (5450), 116-122 (2000)
MEDLINE 20082953
PUBMED 10615043
COMMENT Contact: Vidal M
Marc Vidal laboratory
Dana Farber Cancer Institute
1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
Tel: 617 632 5180
Fax: 617 632 5739
Email: Marc.Vidal@dfci.harvard.edu
Trace dvp10324.x with Bait 11n-53
POLYA=No.

FEATURES
source
1..584
/location/Qualifiers
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/sex="Hermaphrodite and male"
/tissue_type="whole animal"
/dev_stage="mixed stage"
/clone_lib="AD-wrmcDNA"
/note="The AD-wrmcDNA library was generated with poly(A)+
RNA isolated from both hermaphrodite and male N2 worms of
all larval stages, embryos, adults and dauers and the
subsequent generation of cDNAs by poly(A) priming. The
cDNAs were cloned into pCR6"

ORIGIN
Alignment Scores:
Pred. No.: 2,06e-76 Length: 584
Score: 872.00 Matches: 165
Percent Similarity: 100.00% Conservatve: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 59.64% Indels: 0
Gaps: 0

US-09-087-136-1 (1-275) x CB102177 (1-584)

QY 1 Metcerguileapproleuallglupheleuenuprogluapgluapargasna1a 20
Db 88 ATGTCAGAAATGATGCCAGTTCCTGCTCCAGAAAGCGAGATCGAAATGCT 147
QY 21 ArglnaenapproleuileSerglYqlyProleuProleugluserProserArglys 40
Db 148 CGTCAAAATGATCCATTGATAGCGAGGTCCACTTCCTGATGCCCAAGCAAAAA 207
QY 41 LeuthSerleuLeuSerTyraSPProthValProgluSerProaSPMetLysPheA1a 60
Db 208 CTCACATCCGCTGATCTATGATGATCCGACAGTCCGAGTACCGGATATGAAATTCGCC 267
QY 61 ArglysaaggluenglyaaenleuLeuThrTllyshsh1spproserGluileGly 80
Db 268 AGAAAACGTCGGGAAATCTGCTGCAACCAATAAACATCCATCGGAATTAATGGA 327
QY 81 ValneuProgluapTyTThArgAlaapgluugluProglYatrglNglyArgProPro 100

Db 328 GTACTCCAGAAATTAATTAATCTGCTGATGATGAAGACCCGGCCGCAAGACGTCCACA 387
QY 101 G1YarProarglysmetProargHlgIuserSerTherSerleuMetGluSerProarg 120
Db 388 GGTGCCCCGTGAAGATGCGCGTCACGAATCTTCACTTATGGAATTCACACGC 447
QY 121 LyethMetTharGaPserLyel1emecPhegluleuarglyLysProPheglumet 140
Db 448 AAGACTGACTGCTGATTCATTAATTAATGTTGCAATTCGCGTGAACCAATTCGAATG 507
QY 141 l1ealaglyargPhegluugluTyTserleuGlyalargalargVallyGlyH1smet 160
Db 508 ATACTGACGCTTTGAAGAATATTCATCTGTGATGACATGCTTAAAGACACATG 567
QY 161 AspaenGluTyTg1u 165
Db 568 AATTAATGAATATGA 582

RESULT 7
CB102175
LOCUS CB102175 586 bp mRNA linear EST 28-JAN-2003
DEFINITION MW_S0005B10 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA
ACCESSION CB102175
VERSION CB102175.1 GI:27927982
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
AUTHORS Walhout,A.J., Scordella,R., Lu,X., Hartley,J.L., Temple,G.F.,
Brasch,M.A., Thierry-Mieg,N. and Vidal,M.
TITLE Protein interaction mapping in C. elegans using proteins involved
in vulval development
JOURNAL Science 287 (5450), 116-122 (2000)
MEDLINE 20082953
PUBMED 10615043
COMMENT Contact: Vidal M
Marc Vidal laboratory
Dana Farber Cancer Institute
1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
Tel: 617 632 5180
Fax: 617 632 5739
Email: Marc.Vidal@dfci.harvard.edu
Trace dvp10322.x with Bait 11n-53
POLYA=No.

FEATURES
source
1..586
/location/Qualifiers
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/sex="Hermaphrodite and male"
/tissue_type="whole animal"
/dev_stage="mixed stage"
/clone_lib="AD-wrmcDNA"
/note="The AD-wrmcDNA library was generated with poly(A)+
RNA isolated from both hermaphrodite and male N2 worms of
all larval stages, embryos, adults and dauers and the
subsequent generation of cDNAs by poly(A) priming. The
cDNAs were cloned into pCR6"

ORIGIN
Alignment Scores:
Pred. No.: 3.23e-75 Length: 586
Score: 860.00 Matches: 163
Percent Similarity: 100.00% Conservatve: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 58.82% Indels: 0
Gaps: 0

US-09-087-136-1 (1-275) x CB102175 (1-586)

QY 1 MetSerGluIleAspProLeuAlaGluPheLeuLeuProGluAspGlyAspArgAsnAla 20
| | | | |
Db 98 ATCTCAGAAATTAATCACTTCCCGAGCTTTCTTCTCCAGAAACGAGATCCAAATGCT 157
QY 21 ArgGlnAsnAspProLeuIleSerGlyGlyProLeuProLeuGluSerProSerArgLys 40
| | | | |
Db 158 CGTCAAAATGATCATTTGATTAACGAGAGTCCACTTCCATTTGAAATCGCAAGCAGAAAA 217
QY 41 LeuThrSerLeuLeuSerTyrAspProThrValProGluSerProAspMetLysPheAla 60
| | | | |
Db 218 CTACATCCCTGTTATCTTAATGATCCAGACATTCGAGTCCAGATATGAAATTCGCG 277
QY 61 ArgLysArgLeuGlyAsnLeuLeuThrThrIleLysHisHisProSerGluIleIleGly 80
| | | | |
Db 278 AGAAACGCTGTGGAAATCTGCTGACACACATTAACATCCCATCGGAAATTAATTTGA 337
QY 81 ValLeuProGluAspTyrThrArgAlaAspGluGluProGlyArgGlnGlyArgProPro 100
| | | | |
Db 338 GTACTCCAGAAAGATTAATCTCGTCTGATGAAGACCCGGGGCCAGACGATCCACCA 397
QY 101 GlyArgProArgLysMetProArgHisGluSerSerThrSerLeuMetGluSerProArg 120
| | | | |
Db 398 GGTCCGCTCTGTAAGATGCCGCTCAGACATCTTCAACTCTTAATGAAATCACACGCG 457
QY 121 LysThrMetThrArgAspSerLysIleMetPheGluLeuArgGlyLysProPheGluMet 140
| | | | |
Db 458 AAGACTATGACTCGTATCTTAATTAATGTTGAATTCGCGGAAACCATTCGAAATG 517
QY 141 IleAlaGlyArgPheGluGluGluTyrSerLeuGlyArgAlaTyrValLysGlyHisMet 160
| | | | |
Db 518 ATTAGCTGACGCTTTTGAAGAGAAATATTCATTTGTAAGACATGGATTAAAGACACATG 577
QY 161 AsnAsnGlu 163
| | | | |
Db 578 AATTAATGAA 586

RESULT 8
CB102169 562 bp mRNA linear EST 28-JAN-2003
LOCUS MUV_SQ001@B3 AD-wtmcDNA Caenorhabditis elegans cDNA, mRNA sequence.
DEFINITION CB102169
ACCESSION CB102169.1 GI:27927976
VERSION
KEYWORDS
SOURCE EST.
ORGANISM Caenorhabditis elegans
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 562)
Walhout,A.J., Sordella,R., Lu,X., Hartley,J.L., Temple,G.F.,
Brasch,M.A., Thierry-Mieg,N. and Vidal,M.
Protein interaction mapping in C. elegans using proteins involved
in vulval development
JOURNAL Science 287 (5450), 116-122 (2000)
MEDLINE 20082953
PUBMED 10615043
COMMENT Contact: Vidal M
Marc Vidal Laboratory
1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
Tel: 617 632 5180
Fax: 617 632 5739
Email: Marc.Vidal@fci.harvard.edu
Trace dvp10315.x with Bait 1in-53
POLYA=No.

FEATURES
SOURCE Location/Qualifiers
1..562
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/set="Hermaphrodite and male"
/tissue_type="whole animal"

/dev_stage="mixed stage"
/clone_id="AD-wtmcDNA"
/note="The AD-wtmcDNA library was generated with poly(A)+
RNA isolated from both hermaphrodite and male N2 worms of
all larval stages, embryos, adults and dauers and the
subsequent generation of cDNAs by poly(A) priming. The
cDNAs were cloned into pRC6"

ORIGIN

Alignment Scores:
Pred. No.: 6e-74 Length: 562
Score: 847.00 Matches: 160
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.38% Mismatches: 0
Query Match: 57.93% Indels: 0
Gaps: 0

US-09-087-136-1 (1-275) x CB102169 (1-562)

QY 2 SerGluIleAspProLeuAlaGluPheLeuLeuProGluAspGlyAspArgAsnAlaArg 21
| | | | |
Db 79 GCAGAAATTAATCACTTCCCGAGCTTTCTTCTCCAGAAACGAGATCCAAATGCTCGT 138
QY 22 GlnAsnAspProLeuIleSerGlyGlyProLeuProLeuGluSerProSerArgLysLeu 41
| | | | |
Db 139 CAATATGATCATTTGATTAACGAGATCCACTTCCATTTGAAATCGCAAGCAAAATCTC 198
QY 42 ThrSerLeuLeuSerTyrAspProThrValProGluSerProAspMetLysPheAlaArg 61
| | | | |
Db 199 ACATCCCTGTTATCTTAATGATCCAGACATTCGAGTCCAGATCAACGATTAATTCGCAAG 258
QY 62 LysArgLeuGlyAsnLeuLeuThrThrIleLysHisHisProSerGluIleIleGlyVal 81
| | | | |
Db 259 AAACGCTTGGAAATCTGCTGACACACATTAACATCCCATCGGAAATTAATTTGAAGA 318
QY 82 LeuProGluAspTyrThrArgAlaAspGluGluProGlyArgGlnGlyArgProProGly 101
| | | | |
Db 319 CTCACAGAAATTAATCACTTCCCGAGCTTTCTTCTCCAGAAACGAGATCCAAATGCTCG 378
QY 102 ArgProArgLysMetProArgHisGluSerSerThrSerLeuMetGluSerProArgLys 121
| | | | |
Db 379 CGCCCTCTGTAAGATGCCGCTCAGACATCTTCAACTCTTAATGAAATCACACGCAAG 438
QY 122 ThrMetThrArgAspSerLysIleMetPheGluLeuArgGlyLysProPheGluMetIle 141
| | | | |
Db 439 ACTTATGACTCGTATCTTAATTAATGTTGAATTCGCGGAAACCATTCGAAATGATA 498
QY 142 AlaGlyArgPheGluGluGluTyrSerLeuGlyArgAlaTyrValLysGlyHisMetAsn 161
| | | | |
Db 499 GCTGAGACGTTTTGAAGAGAAATATTCATTTGTAAGACATGGATTAAAGACACATGAA 558
QY 162 Asn 162
| | | | |
Db 559 AAT 561

RESULT 9
CB102055 495 bp mRNA linear EST 28-JAN-2003
LOCUS MUV_SQ001@B1 AD-wtmcDNA Caenorhabditis elegans cDNA, mRNA sequence.
DEFINITION CB102055
ACCESSION CB102055.1 GI:27927862
VERSION
KEYWORDS
SOURCE EST.
ORGANISM Caenorhabditis elegans
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 495)
Walhout,A.J., Sordella,R., Lu,X., Hartley,J.L., Temple,G.F.,
Brasch,M.A., Thierry-Mieg,N. and Vidal,M.
Protein interaction mapping in C. elegans using proteins involved
in vulval development
JOURNAL Science 287 (5450), 116-122 (2000)
MEDLINE 20082953

PUBMED 10615043
COMMENT Contact: Vidal M
Marc Vidal Laboratory
Dana Farber Cancer Institute
1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
Tel: 617 632 5180
Fax: 617 632 5739
Email: Marc.Vidal@dfci.harvard.edu
Trace Lin14-AD with Bait Lin-14
POLYA=No.

FEATURES
source Location/Qualifiers
1..495
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/sex="Hermaphrodite and male"
/tissue_type="whole animal"
/dev_stage="mixed stage"
/clone_idb="AD-wrmcDNA"
/note="The AD-wrmcDNA library was generated with poly(A)+ RNA isolated from both hermaphrodite and male N2 worms of all larval stages, embryos, adults and dauers and the subsequent generation of cDNAs by poly(A) priming. The cDNAs were cloned into pPC86"

ORIGIN
Alignment Scores:
Pred. No.: 3,426-61 Length: 495
Score: 718.00 Matches: 137
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 49.11% Indels: 0
DB: 6 Gaps: 0

US-09-087-136-1 (1-275) x CB102055 (1-495)

QY 1 MetSerGluIleAapProLeuAlaGluPheLeuProGluAapGlyAaspArgAana 20
DB 85 ATGTCAGAAATAGATCCACTGCGACGAGTTCTTGCTCCAGAAAGCGAGATCGAAATCT 144
QY 21 ArgGlnAaapProLeuIleSerGlyGlyProLeuProLeuGluSerProSerArglys 40
DB 145 CGTCAAAATGATCCATTATAGCGGAGAGTCCACTTCATTGGAATCCCGAAGCAGAAA 204
QY 41 LeuThrSerLeuLeuSerTyrAapProThrValProGluSerProAaspMetLysPheAla 60
DB 205 CTCACATCCCTGTTATCTATGATCCGACAGTTCGCGAGTCAACCGATATGAAATTCGCC 264
QY 61 ArgGlyAaargLeuGlyAaLeuLeuThrThrIleLysHisIleProSerGluIleIleGly 80
DB 265 AGAAACCTCTGGGAAATCTGCTGCAACCAATAAACATCACCCATCGGAAATATTGGA 324
QY 81 ValLeuProGluAapTyrThrArgAlaAapGluGluProGlyArgGlnGlyArgProPro 100
DB 325 GTACTCCCAAGAAATTAATCTGCTGATGAGAGCCCGGCGCAAGAGATCAACA 384
QY 101 GlyArgProArgLysMetProArgHisGluSerSerThrSerLeuMetGluSerProArg 120
DB 385 GGTGCGCCCTCGTAAGATCCGCGTCACGAATCTTCAACTTCATTATGGAATCAACACGC 444
QY 121 LysThrMetThrArgAapSerLysIleMetPheGluLeuAargGlyLysPro 137
DB 445 AAGACTATGACTCGTGAATTTAAATTTATGTTGAATTGCGTGAACCA 495

RESULT 10
LOCUS CB102055 485 bp mRNA linear EST 28-JAN-2003
DEFINITION MY_S00010B5 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA sequence.
ACCESSION CB102055
VERSION CB102055.1 GI:27927866
KEYWORDS EST.
SOURCE Caenorhabditis elegans

ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida;
Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 485)
AUTHORS Walhout, A.J., Sordella, R., Lu, X., Hartley, J.L., Temple, G.F.,
Brasch, M.A., Thierly-Mieg, N. and Vidal, M.
TITLE Protein interaction mapping in C. elegans using proteins involved
in vulval development
JOURNAL Science 287 (5450), 116-122 (2000)
MEDLINE 20082953
PUBMED 10615043

COMMENT Contact: Vidal M
Marc Vidal Laboratory
Dana Farber Cancer Institute
1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
Tel: 617 632 5180
Fax: 617 632 5739
Email: Marc.Vidal@dfci.harvard.edu
Trace Lin19-AD with Bait Lin-19
POLYA=No.

FEATURES
source Location/Qualifiers
1..485
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/sex="Hermaphrodite and male"
/tissue_type="whole animal"
/dev_stage="mixed stage"
/clone_idb="AD-wrmcDNA"
/note="The AD-wrmcDNA library was generated with poly(A)+ RNA isolated from both hermaphrodite and male N2 worms of all larval stages, embryos, adults and dauers and the subsequent generation of cDNAs by poly(A) priming. The cDNAs were cloned into pPC86"

ORIGIN
Alignment Scores:
Pred. No.: 2,556-58 Length: 485
Score: 689.00 Matches: 131
Percent Similarity: 100.00% Conservative: 3
Best Local Similarity: 97.76% Mismatches: 0
Query Match: 47.13% Indels: 0
DB: 6 Gaps: 0

US-09-087-136-1 (1-275) x CB102059 (1-485)

QY 1 MetSerGluIleAapProLeuAlaGluPheLeuProGluAapGlyAaspArgAana 20
DB 84 ATGTCAGAAATAGATCCACTGCGACGAGTTCTTGCTCCAGAAAGCGAGATCGAAATCT 143
QY 21 ArgGlnAaapProLeuIleSerGlyGlyProLeuProLeuGluSerProSerArglys 40
DB 144 CGTCAAAATGATCCATTATAGCGGAGAGTCCACTTCATTGGAATCCCGAAGCAGAAA 203
QY 41 LeuThrSerLeuLeuSerTyrAapProThrValProGluSerProAaspMetLysPheAla 60
DB 204 CTCACATCCCTGTTATCTATGATCCGACAGTTCGCGAGTCAACCGATATGAAATTCGCC 263
QY 61 ArgGlyAaargLeuGlyAaLeuLeuThrThrIleLysHisIleProSerGluIleIleGly 80
DB 264 AGAAACCTCTGGGAAATCTGCTGCAACCAATAAACATCACCCATCGGAAATATTGGA 323
QY 81 ValLeuProGluAapTyrThrArgAlaAapGluGluProGlyArgGlnGlyArgProPro 100
DB 324 GTACTCCCAAGAAATTAATCTGCTGATGAGAGCCCGGCGCAAGAGATCAACA 383
QY 101 GlyArgProArgLysMetProArgHisGluSerSerThrSerLeuMetGluSerProArg 120
DB 384 GGTGCGCCCTCGTAAGATCCGCGTCACGAATCTTCAACTTCATTATGGAATCAACACGC 443
QY 121 LysThrMetThrArgAapSerLysIleMetPheGluLeuAargGlyLysPro 134

Db 444 AAGACTATGACGTCGATTCATAAATTATGTTGATTGGCT 485

RESULT 11
AV202470 377 bp mRNA linear EST 26-JUL-1999
LOCUS AV202470 Yui Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone yk550h7 5', mRNA sequence.
ACCESSION AV202470 GI:5586241
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Nishigaki, Y., Thierry-Mieg, J., Thierry-Mieg, D., Mitsuiki, H.,
Sano, M., Miyata, A., Motobashi, T., Zeng, Q., Matanabe, H., Sugimoto, A.,
Iida, K., Uesugi, H., Sugiyama, Y. and
Nomoto, H.
Expressed genes in C. elegans
Unpublished (1999)
CONTACT: Yui Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers
1..377
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="yk550h7"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
/clone_lib="Yui Kohara unpublished cDNA"

ORIGIN

Alignment Scores:
Pred. No.: 1,72e-53 Length: 377
Score: 639.00 Matches: 124
Percent Similarity: 99.20% Conservative: 0
Best Local Similarity: 99.20% Mismatches: 0
Query Match: 43.71% Indels: 1
DB: 1 Gaps: 0

US-09-087-136-1 (1-275) x AV202470 (1-377)

QY 9 GIUPELEULEUPROGLUASPGIYASPARGASNAIAARGJINASNAASPPROLEULESER 28
Db 3 GAGTTCCTGCTCCAGAGAAGCGAGATCGAATGCTGCTCAAAATGATCAATTAAGC 62
QY 29 GIYGLYPROLEUPROGLUASPPROSERARGLYLEULETHRSERLEULEUSERTYRAS 48
Db 63 GGAGGTCACCTTCATTGGAATCGCAGCAAGAAATCAATCCCTGTTATCTTATGA 122
QY 48 PPROTHVALPROGLUASPPROASPMETLYSPHEALARGYSARGLEUGIYASNULE 68
Db 123 TCCGACAGTTCGCGAGTCAACCGGATATGAATTCGCCGAAACGCTCGGAAATTCGCT 182
QY 68 UTHTRHLELYSHISHISPROSERGUILLEIGLYVALLEUPROGLUASPTYRTHAR 88
Db 183 GACAAACATAAACAATCAACCATCGGAAATTAATGAGTACTCCCAAGATTATCTCG 242
QY 88 GAAASPGIUGIUPROGLYARGINGIYARGPROGLIYARGPROGLYSMETPROAR 108
Db 243 TCGTCGATGAAGACCCCGGCCCAAGAGCGTCCACAGGTGCCCTGTGAAGATGCCGCG 302
QY 108 GHISGLUSERSERTHRSERLEULETHRSERLEULETHRSERLEULETHRSERLE 128

Db 303 TCACCAATCTTCACTTATGGAATCACCAAGACATGACTCGTGAATCTAA 362

QY 128 silemetphecju 132
Db 363 AATTATGTTTGA 375

RESULT 12
D74918 360 bp mRNA linear EST 18-OCT-1999
LOCUS D74918 CELK090ASFP Yui Kohara unpublished cDNA:Strain N2 hermaphrodite
DEFINITION embryo Caenorhabditis elegans cDNA clone yk90a5 5', mRNA sequence.
ACCESSION D74918 GI:1120702
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Kohara, Y., Mitsuiki, H., Nishigaki, A., Motobashi, T., Sugimoto, A. and
Tabara, H.
Toward an expression map of the C. elegans genome
Unpublished (1994)
CONTACT: Yui Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers
1..360
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk90a5"
/sex="hermaphrodite"
/dev_stage="embryo"
/clone_lib="Yui Kohara unpublished cDNA:Strain N2
hermaphrodite embryo"

ORIGIN

Alignment Scores:
Pred. No.: 1,57e-51 Length: 360
Score: 619.00 Matches: 118
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 42.34% Indels: 0
DB: 7 Gaps: 0

US-09-087-136-1 (1-275) x D74918 (1-360)

QY 1 METSERGUILLEASPPROLEUALAGIUPHELEULEUPROGLUASPGIYASPARGASNAIA 20
Db 6 ATGTCAGAAATGATGATCCAGATTCCTGCTCCAGAGACGAGATGGAATGCT 65
QY 21 ARGGINASNAASPPROLEULESERGIYGLYPROLEUPROLEUGIUSERPROSERARGLYS 40
Db 66 CGTCAAAATGATTCATGTAAGCGAGGTCACCTTCATTTGGAATCCCAAGCAAGAAA 125
QY 41 LEUTHRSERLEULESERTYRASPPTHARVALPROGLUSERPROASPMETLYSPHEAL 60
Db 126 CTCACATCCCTTATTCATGATCCGACAGTCCGCGAGTCAACCGAATTAATTCGCC 185
QY 61 ARGLYSHARGLEUGIYASNULEUTHRTHRLILELYSHISHISPROSERGUILLEIGLY 80
Db 186 AAAAAACGCTGGGAAATCTGTCGACCAACATAAACAATCAACCATCGGAATTAATGA 245
QY 81 VALLEUPROGLUASPTYRTHARVALASPGIUGIUPROGLIYARGINGIYARGPROPRO 100
Db 246 GATCTCCAGAAATTAATCTGCTGATGAAGAGCCCGGCCCAAGAGCGTCCACCA 305

Qy 101 G1YArGProArGlyMeTProArGHisG1uSerThrSerLeuMetG1uSer 118
|
|
|
Db 306 GGTCCCTCGTAAGATCCCGGTACCAATCTTAACCTTCACTTATGGAATCA 359

RESULT 13
LOCUS CB102066
DEFINITION MWY S00010B12 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA
SEQUENCE
ACCESSION CB102066
VERSION CB102066.1 GI:27927873
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditioidea; Rhabditidae; Peloderrinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 438)
AUTHORS Walhout,A.J., Sordella,R., Lu,X., Hartley,J.L., Temple,G.F.,
Brasch,M.A., Thierry-Mieg,N. and Vidal,M.
Protein interaction mapping in C. elegans using proteins involved
in vulval development
JOURNAL Science 287 (5450), 116-122 (2000)
MEDLINE 20082953
PUBMED 10615043
COMMENT Contact: Vidal M
Marc Vidal Laboratory
Dana Farber Cancer Institute
1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
Tel: 617 632 5180
Fax: 617 632 5739
Email: Marc.Vidal@dfci.harvard.edu
Trace Lin8-AD with Bat1 Lin-8
POLYA=No.

FEATURES
source Location/Qualifiers
1..438
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/sex="Hermaphrodite and male"
/tissue_type="whole animal"
/dev_stage="mixed stage"
/clone_lib="AD-wrmcDNA"
/note="The AD-wrmcDNA library was generated with poly(A)+
RNA isolated from both hermaphrodite and male N2 worms of
all larval stages, embryos, adults and dauers and the
subsequent generation of cDNAs by poly(A) priming. The
cDNAs were cloned into pPC86"

ORIGIN
Alignment Scores:
Pred. No.: 2.04e-51 Length: 438
Score: 619.00 Matches: 118
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 42.34% Indels: 0
Gaps: 0
DB: 6
US-09-087-136-1 (1-275) x CB102066 (1-438)

Qy 1 MetSerGluIleAapProLeuAlaGluPheLeuLeuProGluAapGlyAAspArgAsnAla 20
|
|
|
Db 85 ATGTCAAGAAATAGATCCACTTCCGAGTTCTTCTTCAGAAAGACGAGATCGAATCT 144

Qy 21 ArgGlnAsnAapProLeuIleSerGlyGlyProLeuProLeuGluSerProSerArgIlys 40
|
|
|
Db 145 CGTCAAAATGATCATTCATTAAAGCGAGGTCCACTTCATTGGAATGCCCAAGAGAAA 204

Qy 41 LeuThrSerLeuLeuSerTyrAspProThrValProGluSerProAapMetIysPheAla 60
|
|
|
Db 205 CTCACATCCCTGTATATCTATGATCCGACAGTTCGCGAGTCAACCGAATATGAAATCCCC 264

Qy 61 ArgIysArgLeuGlyAAsnLeuLeuThrThrIleIysShiShiProSerGluIleIleGly 80
|
|
|

Db 265 AGAAACGCTCGGGAATCTGCTGACAACTAAAACATCAACCATCGGAATATATGCA 324
|
|
|
Qy 81 ValLeuProGluAapTyrThrArgAlaAapGluGluProGlyArgGlnGlyArgProPro 100
|
|
|
Db 325 GTACTCCCAAGAAATTAATATCTGCTGATATGAAGCCCGGCGCGCAAGAGCTCACCAC 384

Qy 101 G1YArGProArGlyMeTProArGHisG1uSerThrSerLeuMetG1uSer 118
|
|
|
Db 385 GGTCCCTCGTAAGATCCCGGTACCAATCTTAACCTTCACTTATGGAATCA 438

RESULT 14
LOCUS CB102067
DEFINITION MWY S00010C1 AD-wrmcDNA Caenorhabditis elegans cDNA, mRNA sequence.
ACCESSION CB102067
VERSION CB102067.1 GI:27927874
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditioidea; Rhabditidae; Peloderrinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 433)
AUTHORS Walhout,A.J., Sordella,R., Lu,X., Hartley,J.L., Temple,G.F.,
Brasch,M.A., Thierry-Mieg,N. and Vidal,M.
Protein interaction mapping in C. elegans using proteins involved
in vulval development
JOURNAL Science 287 (5450), 116-122 (2000)
MEDLINE 20082953
PUBMED 10615043
COMMENT Contact: Vidal M
Marc Vidal Laboratory
Dana Farber Cancer Institute
1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
Tel: 617 632 5180
Fax: 617 632 5739
Email: Marc.Vidal@dfci.harvard.edu
Trace Lin9-AD with Bat1 Lin-9
POLYA=No.

FEATURES
source Location/Qualifiers
1..433
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/sex="Hermaphrodite and male"
/tissue_type="whole animal"
/dev_stage="mixed stage"
/clone_lib="AD-wrmcDNA"
/note="The AD-wrmcDNA library was generated with poly(A)+
RNA isolated from both hermaphrodite and male N2 worms of
all larval stages, embryos, adults and dauers and the
subsequent generation of cDNAs by poly(A) priming. The
cDNAs were cloned into pPC86"

ORIGIN
Alignment Scores:
Pred. No.: 5.02e-51 Length: 433
Score: 615.00 Matches: 117
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 42.07% Indels: 0
Gaps: 0
DB: 6
US-09-087-136-1 (1-275) x CB102067 (1-433)

Qy 1 MetSerGluIleAapProLeuAlaGluPheLeuLeuProGluAapGlyAAspArgAsnAla 20
|
|
|
Db 81 ATGTCAAGAAATAGATCCACTTCCGAGTTCTTCTTCAGAAAGACGAGATCGAATCT 140

Qy 21 ArgGlnAsnAapProLeuIleSerGlyGlyProLeuProLeuGluSerProSerArgIlys 40
|
|
|
Db 141 CGTCAAAATGATCATTCATTAAAGCGAGGTCCACTTCATTGGAATGCCCAAGAGAAA 200

QY 41 LeuThrSerLeuLeuSerTyrAspProThrValProGluSerProAspMetLysPheAla 60
 Db 201 CTCACATCCCTGTTATCTATGATCCGACAGTCCGAGTCACCGGATATGAATTCGCC 260
 QY 61 ArgLysArgLeuGlyAsnLeuLeuThrThrIleLysHisProSerGluIleIleGly 80
 Db 261 AGAAACGCTCGGGAATCTGCTGACAAACCATTAACATCAACCATCGAAATATATGGA 320
 QY 81 ValLeuProGluAspTyrThrArgAlaAspGluGluProGlyArgGlnGlyArgProPro 100
 Db 321 GTACTCCCAAGAAATTATACCTGCTGATGAAGAGCCGCGGCCCAAGACGTCACCA 380
 QY 101 GlyArgProArgLysMetProArgHisGluSerSerThrSerLeuMetGlu 117
 Db 381 GGTGGCCCTCGTAAGATGCCGCGTACGAACTTCAACTTCATTATGAA 431

RESULT 15

LOCUS CB102064 425 bp mRNA linear EST 28-JAN-2003
 DEFINITION MIM S0001@B10 AD-wrmcDNA Caenorhabditis elegans CDNA, mRNA
 sequence.

ACCESSION

CB102064

CB102064.1

GI:27927871

EST.

Caenorhabditis elegans

Caenorhabditis elegans

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.Walhout, A.J., Sordeja, R., Lu, X., Hartley, J.L., Temple, G.F.,
 Brasch, M.A., Thierly-Mieg, N. and Vidal, M.
 Protein interaction mapping in C. elegans using proteins involved
 in vulval development
 Science 287 (5450), 116-122 (2000)

JOURNAL

MEDLINE

20082953

10615043

PUBMED

COMMENT

Contact: Vidal M
 Dana Farber Cancer Institute
 1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
 Tel: 617 632 5180
 Fax: 617 632 5739
 Email: Marc.Vidal@dfci.harvard.edu
 Trace Line AD with Bait 1in-6
 POLYA=No.

FEATURES

source

Location/Qualifiers

1..425

/organism="Caenorhabditis elegans"

/mol_type="mRNA"

/strain="N2"

/db_xref="taxon:6239"

/sex="Hermaphrodite and male"

/tissue_type="whole animal"

/dev_stage="mixed stage"

/clone_lib="AD-wrmcDNA"

/note="The AD-wrmcDNA library was generated with poly(A)+
 all larval stages, embryos, adults and dauers of the
 subsequent generation of CDNA by poly(A) priming. The
 CDNA were cloned into pC86"

ORIGIN

Alignment Scores:

Pred. No.: 4.83e-50 Length: 425

Score: 605.00 Matches: 115

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 41.38% Indels: 0

DB: 6 Gaps: 0

US-09-087-136-1 (1-275) x CB102064 (1-425)

QY 1 MetSerGluIleAspProLeuAlaGluPheLeuLeuProGluAspGlyAspArgAsnAla 20
 Db 79 ATGTGAAATATGATTCCTGATGATCCGAGTCTTGCTCCAGAAAGCGGATGGAATGCT 138
 QY 21 ArgGlnAsnAspProLeuIleSerGlyGlyProLeuProLeuGluSerProSerArgLys 40
 Db 139 CGTCAAAATGATCCATTGATGAAGCGAGGTCCATTCCATTGGAAATGCCAAGCAAAAA 198
 QY 41 LeuThrSerLeuLeuSerTyrAspProThrValProGluSerProAspMetLysPheAla 60
 Db 199 CTCACATCCCTGTTATCTATGATCCGACAGTCCGAGTCACCGGATATGAATTCGCC 258
 QY 61 ArgLysArgLeuGlyAsnLeuLeuThrThrIleLysHisProSerGluIleIleGly 80
 Db 259 AGAAACGCTCGGGAATCTGCTGACAAACCATTAACATCAACCATCGAAATATATGGA 318
 QY 81 ValLeuProGluAspTyrThrArgAlaAspGluGluProGlyArgGlnGlyArgProPro 100
 Db 319 GTACTCCCAAGAAATTATACCTGCTGATGAAGAGCCGCGGCCCAAGACGTCACCA 378
 QY 101 GlyArgProArgLysMetProArgHisGluSerSerThrSerLeu 115
 Db 379 GGTGGCCCTCGTAAGATGCCGCGTACGAACTTCAACTTCATTATGAA 423

Search completed: April 9, 2005, 10:33:59
 Job time : 3516 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 9, 2005, 09:40:53 ; Search time 624 Seconds

(Without alignments)
2670.675 Million cell updates/sec

Title: US-09-087-136-1

Perfect score: 1462
Sequence: 1 MSERDPLAFLLPFDGDRNA.....INKYRGESKTWVQALGLI 275

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 5615251 seqs, 3030001701 residues

Total number of hits satisfying chosen parameters: 11230502

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPRO/spool/US09087136/runat_07042005_163618_1039/app_query_faasta_1.455
-DB=Published Applications NA -OFMT=fastlap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPEXT=0 -INITS=bits -STRAT=1 -END=1 -MATRIX=blowum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTEXT=pct -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09087136_@CGN1_1_698_@runat_07042005_163618_1039
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DISPATCH=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA: *
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq: *
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq: *
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq: *
4: /cgn2_6/ptodata/1/pubpna/US05_PUBCOMB.seq: *
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq: *
6: /cgn2_6/ptodata/1/pubpna/PCTOS_PUBCOMB.seq: *
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq: *
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq: *
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq: *
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq: *
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq: *
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq: *
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq: *
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq: *
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq: *
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq: *
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq: *
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq: *
19: /cgn2_6/ptodata/1/pubpna/US10F_NEW_PUB.seq: *
20: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq: *
21: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq: *
22: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq: *

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1462	100.0	954	US-09-220-091-2	Sequence 2, Appl1
2	109.5	7.5	1592	US-10-437-963-31520	Sequence 31520, A
3	109	7.5	1335	US-10-739-930-288	Sequence 288, App
4	107.5	7.4	1428	US-10-363-345A-1791	Sequence 1791, Ap
5	107.5	7.4	1428	US-10-363-345A-1792	Sequence 1792, Ap
6	107.5	7.4	1428	US-10-363-483A-1791	Sequence 1791, Ap
7	107.5	7.4	1428	US-10-363-483A-1792	Sequence 1792, Ap
8	107.5	7.4	1428	US-10-363-483A-1791	Sequence 1791, Ap
9	107.5	7.4	1428	US-10-363-483A-1792	Sequence 1792, Ap
10	107	7.3	9369	US-10-200-562-190	Sequence 190, Appl
11	107	7.3	9369	US-10-237-551-247	Sequence 247, App
12	107	7.3	9369	US-10-237-551-247	Sequence 247, App
13	106	7.3	1590	US-10-156-761-2032	Sequence 2032, Ap
14	106	7.3	1590	US-10-156-761-2032	Sequence 2032, Ap
15	105	7.2	1041	US-10-156-761-5319	Sequence 5319, Ap
16	105	7.2	1041	US-10-156-761-5319	Sequence 5319, Ap
17	104.5	7.1	1792	US-10-425-114-15465	Sequence 15465, A
18	104	7.1	954	US-10-425-114-4732	Sequence 4732, Ap
19	104	7.1	1793	US-10-739-930-2962	Sequence 2962, Ap
20	104	7.1	3183	US-10-425-115-96511	Sequence 96511, A
21	104	7.1	3289	US-09-963-285-8	Sequence 8, Appl1
22	104	7.1	3289	US-10-806-782-1	Sequence 1, Appl1
23	104	7.1	6458	US-09-963-285-1	Sequence 1, Appl1
24	103.5	7.1	6491	US-09-954-456-2118	Sequence 2118, Ap
25	103.5	7.1	6491	US-10-172-118-559	Sequence 559, App
26	103.5	7.1	6491	US-10-342-887-559	Sequence 559, App
27	103.5	7.1	6491	US-10-648-593-84	Sequence 84, Appl
28	103.5	7.1	6491	US-10-648-593-84	Sequence 84, Appl
29	103.5	7.1	6608	US-10-843-641A-5145	Sequence 5145, Ap
30	102	7.0	2160	US-10-116-802-15	Sequence 15, Appl
31	102	7.0	101365	US-10-156-761-6853	Sequence 6853, Ap
32	101.5	6.9	939	US-10-719-993-6830	Sequence 6830, Ap
33	101.5	6.9	939	US-10-363-345A-27128	Sequence 27128, A
34	101.5	6.9	939	US-10-363-345A-27128	Sequence 27128, A
35	101.5	6.9	939	US-10-363-483A-27128	Sequence 27128, A
36	101.5	6.9	939	US-10-363-483A-27128	Sequence 27128, A
37	101.5	6.9	1095	US-10-363-483A-27128	Sequence 3625, A
38	101.5	6.9	1596	US-10-425-115-88283	Sequence 88283, A
39	101.5	6.9	1640	US-10-425-115-32254	Sequence 32254, A
40	101.5	6.9	1645	US-10-425-114-30266	Sequence 30266, A
41	101.5	6.9	1803	US-10-425-115-133999	Sequence 133999, A
42	101.5	6.9	2775	US-10-359-493-42070	Sequence 42070, A
43	100.5	6.9	1463	US-10-437-963-98897	Sequence 98897, A
44	100.5	6.9	1201	US-10-017-161-2025	Sequence 2025, Ap
45	100.5	6.9	4505	US-10-292-798-1671	Sequence 1671, Ap
				US-10-437-963-5126	Sequence 5126, Ap

ALIGNMENTS

RESULT 1
US-09-220-091-2
Sequence 2, Application US/09220091
Patent No. US20020064523A1
GENERAL INFORMATION:
APPLICANT: H. Robert Horvitz
APPLICANT: Craig Ceol
APPLICANT: Xiaowei Lu
TITLE OF INVENTION: A TUMOR SUPPRESSOR PATHWAY IN C. ELEGANS
FILE REFERENCE: 0197/202003
CURRENT FILING DATE: 1998-12-23
EARLIER APPLICATION NUMBER: 60/047,996
EARLIER FILING DATE: 1997-05-28
EARLIER APPLICATION NUMBER: 09/087,136
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 954

```

; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
US-09-220-091-2

Alignment Scores:
Pred. No.: 9,186-159 Length: 954
Score: 1462.00 Matches: 275
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-087-136-1 (1-275) x US-09-220-091-2 (1-954)

QY 1 MetSerGluIleAspProLeuAlaGluPheLeuLeuProGluAspGlyAspArgAsnAla 20
DB 1 ARGTCAGAAATGATCCACTTGCACGATTCCTGCTCCAGAGAGCGAGATCGAAATGCT 60
QY 21 ArgGlnAsnAspProLeuIleSerGlyGlyProLeuProLeuGluSerProSerArgLys 40
DB 61 COTCAAAATGATCCATTGATTAAGCGAGGTCCACTTCCATTGGAAATCCGCAAGCAGAAA 120
QY 41 LeuThrSerLeuLeuSerLysThrAspProThrValProGluSerProAspMetLysPheAla 60
DB 121 CTCACATCCCTGTATCTATCTATGATCCGACAGTCCGAGGTCCAGCATATGAAATTCGCC 180
QY 61 ArgLysArgLeuGlyAsnLeuLeuThrTrpIleLysHisProSerGluIleIleGly 80
DB 181 AGAAAACGTGGGAAATCTGTGCAACCAATMAAACATCACCCATCGGAAATATTTGGA 240
QY 81 ValLeuProGluAspLysThrArgAlaAspGluGluProGlyArgGlnGlyArgProPro 100
DB 241 GTACTCCCAAGAAAGATTATATCTGTCTGATGAAAGCCCGGCGGCAAGAGCTCACCA 300
QY 101 GlyArgProArgLysMetProArgHisGluSerSerThrSerLeuMetGluSerProArg 120
DB 301 GGTCCCTCCCTGTAAGATGCCGCGTCCAGCATCTTCACTTATGGAATCACACACGC 360
QY 121 LysThrMetThrArgAspSerLysIleMetPheGluLeuArgGlyLysProPheGluMet 140
DB 361 AAGACTTAAGACTCGGATTTCTMAAATTTATGTTTGAAATGGCGGAAAACCATTCGAAATG 420
QY 141 IleAlaGlyArgPheGluGluGluLysSerLeuGlyArgAlaTrpValLysGlyHisMet 160
DB 421 ATAGCTGGACGTTTTGAAAGAAATATTCACCTGTGTAAGCATGTGTTAAAGGACATG 480
QY 161 AsnAsnGluLysGluProIleLysAlaGlnArgThrAspTyrAlaProAsnLeuAlaVal 180
DB 481 AATAATGAAATGAAACCAATMAAGCTCAAGAGACAGACTATGCAACCAATCTGCTGTT 540
QY 181 AspTyrLeuAlaCysArgGluIleHisArgMetProArgProAspLysSerIleProGlu 200
DB 541 GATTATCTTGGCATGTGCGAGATTCATGGAATGCCAGCTCCAGATTAATTCATTTCTGAG 600
QY 201 LeuProIleValProSerArgIleAspGluPheAspAlaThrValAspProArgTyrGlu 220
DB 601 CTGCCCAATTGTTCCATCTAGATCGATGATTCGACGTACAGTCGATCAAGATATGAA 660
QY 221 ThrAspLeuLysAsnGluLysTrpIleArgHisTrpLysGlnValLysGlyTrpCysAla 240
DB 661 ACAGATTTGAAATAAGAAATACATTTGTCATTTGGAACAGTCMAAAAGGTTGGGTGCT 720
QY 241 HisGlnArgArgArgThrAlaProHisAlaArgSerIleAlaLeuIleAsnLysIleLys 260
DB 721 CATCAAGCTGCTGAGCTGCTCCCATGCAAGACAGACATGACATTAATCAAAATCTTAC 780
QY 261 GlnProGlyGluSerLysThrValGluGlnAlaLeuGlyLeuIle 275
DB 781 CAGCTGGAAGAGTCAAAAACGTCTCGAGCAAGCACTTGTGTTATT 825

RESULT 2
US-10-437-963-31520
; Sequence 31520, Application US/10437963
```

```

; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21 (53221)B
; CURRENT APPLICATION NUMBER: US/10/437, 963
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 31520
; LENGTH: 1592
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_35819C.1
US-10-437-963-31520

Alignment Scores:
Pred. No.: 0,0279 Length: 1592
Score: 109.50 Matches: 59
Percent Similarity: 36.45% Conservative: 19
Best Local Similarity: 27.57% Mismatches: 90
Query Match: 7.49% Indels: 46
DB: 18 Gaps: 10

US-09-087-136-1 (1-275) x US-10-437-963-31520 (1-1592)

QY 5 AspProLeuAlaGluPheLeuProGluAspGlyAspArgAsnAlaArgGlnAsnAsp 24
DB 356 GATCCTCTTCGACGAGACCCCTCTACAGAAAGCCAGAGATGCAAGCCCTTCGTCATGT 415
QY 25 ProLeuIleSerGlyGlyProLeuProLeuGluSerProSerArgLysLeuThrSerLeu 44
DB 416 CCT-----CAGAGAGGCGCGTCCC-----TCCCGGCGATAAAGTGAAGGCA 463
QY 45 LeuSerTyrAspProThrValProGluSerProAspMetLysPheAlaArgLysArgLeu 64
DB 464 CCATCGAGGTGCGCGGACCGACGAAAGAACACACCAAGGCGCAGATGACTCGGCA 523
QY 65 GlyAsnLeuLeuThrTrpIleLysHisHisProSerGluIleIleGlyVal-LeuProGlu 84
DB 524 AGCAATGCGCCCAAGTACTTACAGAGCGGTGCCCTTCGCCMAAGTGCCTGCTCA 583
QY 84 uAspTyrThrArgAla-----AspGluGluProGly----- 94
DB 584 AGATCGAGCCCAACCAAGCCCTCCAGCTTGCCATTGATCTGAAGCCCAAGGCTTGCTT 643
QY 95 -----ArgGlnGly-----ArgProProGlyArgProAr 104
DB 644 GCTACGCATCATCTGCGCAGAGAAACGGGCTGGGCGCGATCGTGAAGCGGAGATCTCG 703
QY 104 GlyMetPro---ArgHisGluSer---SerThrSerLeuMetGluSerProArgLysTrp 122
DB 704 TCGATGCGCCCTCACGACATCATGATCGAGCGGATCGTCCGAGAGTGTCTTCCGCCGT 763
QY 122 rMetThrArgAspSerLysIleMetPheGluLeuArgGlyLysProPheGluMetIleAl 142
DB 764 GCTTCAAGGCGCTGAAGAGACCATCTCTCTCGAGGCGACCTCTTAAGGCCCAACA 823
QY 142 aglyArg-----PheGluGluGluLysSerLeuGlyArgAlaTrpValLysGlyHis 159
DB 824 TGGTCAACCCCGGCTTCGACGCGCAAGAAAGTGGCGCGGA-----GGTCA 868
QY 159 eMetAsnAsnGluLysGluProIleLysAlaGlnArgThrAspTyrAlaProAsnLeuAl 179
DB 869 T-----CGCGAGATCACCGTCCGACCCCTCC 895
```

```
Oy 179 aValAspTyrLeuAlaCysArgGluIleHisArgMetPro 192
Db 896 AGAGGACCGCTCCCTCCGCCGCTGCCGACATGCTTCT 935

RESULT 3
US-10-739-930-288
; Sequence 288, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 288
; LENGTH: 1335
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: clone ID: ARATH-23APR03-CLUSTER15254_1
US-10-739-930-288

Alignment Scores:
Pred. No.: 0.0254 Length: 1335
Score: 109.00 Matches: 69
Percent Similarity: 36.93% Conservative: 37
Best Local Similarity: 24.04% Mismatches: 99
Query Match: 7.46% Indels: 82
Gaps: 16

US-09-087-136-1 (1-275) x US-10-739-930-288 (1-1335)
Oy 5 AspPro-----LeuAlaGluPheLeuLeuProGluAspArgAsnAlaArg 21
Db 321 GATCGCGACGCTGCTTACCCGTTGCTTACCGGCTCAGGG---TTTTCAGCCCGA 377
Oy 22 Gln-----AsnAspProLeuIleSerGlyIleProLeuProLeuGluSer 36
Db 378 CCTGTTCGAGGATTCGTCGACGATCTTCCTCGACGCGTGAATCTGAGTGGTACCCT 437
Oy 37 ProSerArgLysLeuThrSerLeuLeuSerTyrAspProThrValProGluSerProAsp 56
Db 438 CCTGCCCGGAGTTTACT-----TACGATCTCGAGCAATATGACACGCTCAA 485
Oy 57 Met-----LysPheAlaArgLysArgLeuGlyAsnLeuLeuThrThrIle 71
Db 486 ATGGAATCCTTGTTCGACGACATTTATAGGAGAGAAATCTCTCAGATT----- 533
Oy 72 LysHisHisProSerGluIleIleGlyValLeuProGluAspTyrThrArgAlaAspGlu 91
Db 534 -----CGTCCATTACCAACGCGCTTGAGATTG----- 557
Oy 92 GluProGluArgGlnIleArgProProGly----- 101
Db 558 -----GGTTCTCTGTTGGGTGGTGGTCCCATTTAGACAAAGTCTCAATT 602
Oy 102 ---ArgProArgLysMetProArgHisGluSerSerThrSerLeuMetGluSerPro--- 119
Db 603 CTGCAACCGCGGCTTGGCCCG-----CCCCCACTTCATTTCTAGACACTAGCAGG 653
Oy 120 ---ArgLysThrMetThrArgAspSerLysIleMetPheGluLeuArgGlyLysProPhe 138
Db 654 AATGAAAGGCCAAGACAGAGATGGGGCTCTCCGTT---GTTAGGGGAGAAAGATC 710
Oy 139 GluMetIleAlaGlyArgPheGluGluGluSerLeuGlyArgAlaThrValIleGly 158
Db 711 CGAATCTACTGAGGGCTCT---TCTTCTTTATTCATCTGTCGATCGGTGAAA--- 764
Oy 159 HisMetAsnAsnGluTyrGluProIleLysAlaGlnArgThrArgAlaProAsnLeu 178
```

```
Db 765 -----AATGCTGCTCAGTTGGATTCACGCCGCAAGAGATGGTATA----- 806
Oy 179 AlavalAspTyrLeuAlaCysArgGluIleHisArgMetProArgPro----- 194
Db 807 -----ATGAACCTTTGGCCAAACCGTTACTGTGAT 839
Oy 195 -----AspLysSerIleProGluLeuProIleValProSerArgIleAsp----- 209
Db 840 TTAAACAAGACAGACATGTACCGATGATCAGATGACATGAAGATCAGCGATGAACAAA 899
Oy 210 GluPheAspAlaThrValAspProArgTyrGluThrAspLeuLysAsnGluTyrIleArg 229
Db 900 GAGGACGAGAGAGCTGTAAAGCAATATCTGAGAGAGATCTTTGAAGAAGACATCGAG 959
Oy 230 HisTrpLysGlnValLysLysGlyTyrCysAlaHisGlnArgArgThrAlaProHis 249
Db 960 CGAGCTAAGAAAGTCCGTGCACAAATTACGAGAAAGCGTCGAGAGATCAGAGATAC 1019
Oy 250 AlaArgSerIleAlaLeuIle 256
Db 1020 AAAGAAAGAAATTAATCTTATT 1040

RESULT 4
US-10-363-345A-1791/c
; Sequence 1791, Application US/10363345A
; Publication No. US20040234960A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Method for determining the degree of methylation of defined
; FILE REFERENCE: E01/1227
; CURRENT APPLICATION NUMBER: US/10/363,345A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 1791
; LENGTH: 1428
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-363-345A-1791

Alignment Scores:
Pred. No.: 0.0412 Length: 1428
Score: 107.50 Matches: 74
Percent Similarity: 35.33% Conservative: 38
Best Local Similarity: 23.34% Mismatches: 113
Query Match: 7.35% Indels: 93
Gaps: 15

US-09-087-136-1 (1-275) x US-10-363-345A-1791 (1-1428)
Oy 12 LeuProGluAspGlyAspArgAsnAlaArgGlnAsnAspProLeuIleSerGlyIlePro 31
Db 1247 CTACCGCAGCAAAACCCCGCAAGCGTAAACGCCG-AATCCGCGCTCCCTCCGCGCG 1189
Oy 32 LeuPro-----LeuGluSerProSerArg----- 39
Db 1188 ACCCCTAACCGCGCCCGCAAAATTAACCGACCTACCAACTTCAAAAAACCGCCCGCACTCA 1129
Oy 40 -----LysLeuThrSerLeuLeuSer-TyrAspProThrValProGluSerPr 55
Db 1128 AAACGCGAGATCCGCCGACGCGCTCGACCTCGACACTCAACATTAATCC----- 1077
Oy 55 cAspMetLysPheAlaArgLysArgLeuGlyAsnLeuLeuThrThrIleLysHisHisPr 75
Db 1076 -----GCGGCGCCTCGACGACAAACCTTCGCAAAAGCCCTTAACAAAACCC 1030
Oy 75 cSerGluIleIleGlyValLeuProGluAspTyrThrArgAlaAspGluGluProGluYar 95
```

```

Db      1029 GACCC-----CCGACCCCGACCTGACCTCGGACGACAAACCC----- 993
Qy      95 gGInGlyArgProGlyArgProArglySmetProArgHISgluSerSerTherSeri 115
Db      992 -----GAACCCCGCCGAAACCGACCGACGACGCGAATAAATAATCGTACGACGA 940
Qy      115 uMetGluSerProArg-----LysThrMetThrArgAsp----- 126
Db      939 AAAAAAAAAATGCAAAAAACAACAATCCGAAATTCMAAAGAAAAACCGACCGAATAA 880
Qy      127 -----SerLysIleMetPheGluLeuArgGly 136
Db      879 AAAAAAGACGCGCCCGCGCCCTTTTATACAAAAAATCCGAACGAACCTCGCGCCCG 820
Qy      136 sPro-----PheGluMetIleAlaGlyArgPheGlu----- 146
Db      819 CCGACCGCACTTTCAACGCAAACTCTATTTCCAAACATCACTCGACGACCAAAATATCC 760
Qy      147 -----GluGluTyrSerLeuGlyArgAlaTrpValLysGlyHISMetAs 161
Db      759 TCGGAACGCTTCCAAACAAACTACGCTATAATTAACTATACACTTAAAAAACGGCA 700
Qy      161 pAsnGluTyrGluProIleLysAlaGlnArgThrAspTyrAlaProAsnLeuAlaValAs 181
Db      699 CCGAACGCAAGATCTTAAAAAATAAATACGCGCCGACCGCGCCGAAACCGCCCGCG-- 642
Qy      181 pTyrLeuAlaCybArgGluIleHISArgMetProArgProArgLysSerIlePro----- 199
Db      641 -----CAACCCCGCCACCGCTCGCGAACAACCGCCCGCGCGCCCTCC 595
Qy      200 -----GluLeuProIleValProSerArgIleAspGluPh 211
Db      594 AAACCCGACCTCGGCTAACGCAACCCCTCCG---CCCCCGGCGCCACGCGCGACT 538
Qy      211 eAspAlaThrValAspProArgTyrGluThrAspLeuLysAsnGluTyrIleArgHISTr 231
Db      537 CGACGCGACCC---TCGCCCCCTATTAATAAATAATCTTTAAACGATCCGCTAAAA 481
Qy      231 pLysGln-----ValLysGlyTrpCysAlaHISgluArgAspArgTh 246
Db      480 CCGAACTTCACCCCACTTTTACAAACAAACCTTCGACGACGCTCGACGCGCGAAC 421
Qy      246 rAlaProHISAlaArgSerIleAlaLeuIleAsnLysIleTyrGlnPro 262
Db      420 GACCCCG-----AAATCCATGCTCTTAAAAAATATTATTATTTCCCG 378

RESULT 5
US-10-363-345A-1792
/ Sequence 1792, Application US/10363345A
/ Publication No. US20040234960A1
/ GENERAL INFORMATION:
/ APPLICANT: Alexander Olek
/ APPLICANT: Christian Piepenbrock
/ APPLICANT: Kurt Berlin
/ TITLE OF INVENTION: Method for determining the degree of methylation of defined
/ FILE REFERENCE: E01/1227
/ CURRENT APPLICATION NUMBER: US/10/363,345A
/ NUMBER OF SEQ ID NOS: 40712
/ SEQ ID NO 1792
/ LENGTH: 1428
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-363-345A-1792

Alignment Scores:
Pred. No.: 0.0412 Length: 1428
Score: 107.50 Matches: 74
Percent Similarity: 35.33% Conservative: 38

```

```

Best Local Similarity: 23.34% Mismatches: 113
Query Match: 7.35% Indels: 93
DB: 18 Gaps: 15

US-09-087-136-1 (1-275) x US-10-363-345A-1792 (1-1428)

Qy      12 LeuProGluAspArgAlaAsnAlaArgGlnAsnAspProLeuIleSerGlyPro 31
Db      182 CTACCCGACCAACCCCCCGAAACGTTAAACGCCG -AATCCGCGCTCCCTCCGCGCG 240
Qy      32 LeuPro-----LeuGluSerProSerArg----- 39
Db      241 ACCCTTAACGCGCCCGGAATTAACGCACTTACCAATTCGAAAAAACCGCCGACATCA 300
Qy      40 -----LysLeuThrSerLeuLeuSer-TyrAspProValProGluSerPr 55
Db      301 AAACGCGAATCCCGCGACGCGCTCGACCTCGACCTGCACTGCACTGCACTTAATCC 352
Qy      55 oAspMetLysPheAlaArgLysArgLeuGlyAsnLeuLeuThrIleLysHISISPr 75
Db      353 -----GCGGCGCTGACGACGAAACCTGCAAAACGCGCCCTTAACAAACCC 399
Qy      75 oSerGluIleIleGlyValLeuProGluAspTyrThrArgAlaAspGluGluProGlyAr 95
Db      400 GACC-----CCGACCCCGACCTGACCTCGGACGACAAACCC----- 436
Qy      95 gGInGlyArgProGlyArgProArglySmetProArgHISgluSerSerTherSeri 115
Db      437 -----GAACCCCGCCGAAACCGACCGACGCTCGAATAAATAATCGTACGACGA 489
Qy      115 uMetGluSerProArg-----LysThrMetThrArgAsp----- 126
Db      490 AAAAAAAAAATGCAAAAAACAACAATCCGAAATTCMAAAGAAAAACCGACGAAATAA 549
Qy      127 -----SerLysIleMetPheGluLeuArgGly 136
Db      550 AAACACGACGCGCCCGCGCCCTTTTATACAAAAAATCGAAGAACTCCGCGCCCG 609
Qy      136 sPro-----PheGluMetIleAlaGlyArgPheGlu----- 146
Db      610 CCGACCGCACTTTCAACGCAAACTCTATTTCCAAACATCACTGACGACGCAAAATATCC 669
Qy      147 -----GluGluTyrSerLeuGlyArgAlaTrpValLysGlyHISMetAs 161
Db      670 TCGGAACGCTTCCAAACAAACTACGCTATAATTAACTATACACTTAAAAAACGGCA 729
Qy      161 pAsnGluTyrGluProIleLysAlaGlnArgThrAspTyrAlaProAsnLeuAlaValAs 181
Db      730 CCGAACGCAAGATCTTAAAAAATAAATACGCGCCGACCGCGCCGAAACCGCCCGCG-- 787
Qy      181 pTyrLeuAlaCybArgGluIleHISArgMetProArgProArgLysSerIlePro----- 199
Db      788 -----CAACCCCGCCACCGCTCGCGAACAACCGCCCGCGCGCCCTCC 834
Qy      200 -----GluLeuProIleValProSerArgIleAspGluPh 211
Db      835 AAACCGGACCTCGGCTAACGCAACCCCTCCG---CCCCCGGCGCCACGCGCGACT 891
Qy      211 eAspAlaThrValAspProArgTyrGluThrAspLeuLysAsnGluTyrIleArgHISTr 231
Db      892 CGACGCGACCC---TCGCCCCCTATTAATAAATAATCTTTTAAACGATCCGCTAAAA 948
Qy      231 pLysGln-----ValLysGlyTrpCysAlaHISgluArgAspArgTh 246
Db      949 CCGAACTTCACCCCACTTTTACAAACAAACCTTCGACGACGCTCGACGCGCGAAC 1008
Qy      246 rAlaProHISAlaArgSerIleAlaLeuIleAsnLysIleTyrGlnPro 262
Db      1009 GACCCCG-----AAATCCATGCTCTTAAAAAATATTATTATTTCCCG 1051

RESULT 6
US-10-363-483A-1791/C
/ Sequence 1791, Application US/10363483A

```

```
Publication No. US20050064401A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
; FILE OF INVENTION: 11111111
; FILE REFERENCE: 82011
; CURRENT APPLICATION NUMBER: US/10/363,483A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 1791
; LENGTH: 1428
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: CPG-Island No: 1791
US-10-363-483A-1791

Alignment Scores:
Pred. No.: 0.0412 Length: 1428
Score: 107.50 Matches: 74
Percent Similarity: 35.33% Conservative: 38
Best Local Similarity: 23.34% Mismatches: 113
Query Match: 7.35% Indels: 93
DB: 19 Gaps: 15

US-09-087-136-1 (1-275) x US-10-363-483A-1791 (1-1428)
Qy 12 LeuProGluAerGlyAspArgAsnAlaArgGlnAsnAspProLeuIleSerGlyGlyPro 31
Db 1247 CTACCCGACCAAAACCCCGGAAAGCTTAAGCGCG-AATCCGCGCTCCCTCCGCGCG 1189
Qy 32 LeuPro-----LeuGluSerProSerArg----- 39
Db 1188 ACCCTTAACGCGCCCGAAATTAACGACCTTAACCACTTAACCAAAACCCCGCGACATCA 1129
Qy 40 -----LysLeuThrSerLeuLeuSer-TyrAspProThrValProGluSerPr 55
Db 1128 AAACGCGGATCCGCGCGACCGCTCGACCTCGACATCATTAAATCC----- 1077
Qy 55 oAspMetLysPheAlaArgLysArgLeuGlyAsnLeuLeuThrThrIleLysHisLysPr 75
Db 1076 -----GCGCGCGCTCGACGACGAAACCTCGCAAAACGCCCTTAACAAACCC 1030
Qy 75 oSerGluIleIleGlyValLeuProGluAspTyrThrArgAlaAspGluGluProGlyAr 95
Db 1029 GACC-----CCGACCCCGGACTCGCGCGCGACCAACCC----- 993
Qy 95 gGlnGlyArgProGluArgProGluArgLysMetProArgHisGluSerSerThrSerLe 115
Db 992 -----GAACCCCCCGAACCGACCGGACGCTGAAATAAATAATGTACGACGA 940
Qy 115 uMetGluSerProArg-----LysThrMetThrArgAsp----- 126
Db 939 AAAAAAAAAATCGAAAAACAATCCGGAATTCAAAAAGAAAAACGACCGCAAAAAA 880
Qy 127 -----SerLysIleMetPheGluLeuArgGlyLys 136
Db 879 AAAAAGACGCGCGCGCGCGCTTTATACAAATAATCGAAGAACCTCCGCGCGCG 820
Qy 136 sPro-----PheGluMetIleAlaGlyArgPheGlu----- 146
Db 819 CCGACCGGACTTTCACGCAAAATCTTAATTCACCAATCACTCGACGACCAATATCC 760
Qy 147 -----GluGluTyrSerLeuGlyArgAlaTrpValLysGlyHisMetLys 161
Db 759 TCCGACGCTTCGCAAAACAATACGTAATAATTATACATTAACCTTCAAAAAACGCGAA 700
Qy 161 nAsnGluTyrGluProIleLysAlaGlnArgThrAspTyrAlaProAsnLeuAlaValAs 181
Db 699 CCGAAGCAGCATCTTAATAAATAAATAATACGCGCGCGCGCGCGCGCGCGCGCG-- 642
```

```
Qy 181 pTyrLeuAlaCyArgGluIleHisArgMetProArgProAspLysSerIlePro----- 199
Db 641 -----CAACCCGCGCGCGCTCGGAAACCGCGCGCGCGCGCGCGCGCGCGCG 595
Qy 200 -----GluLeuProIleValProSerArgIleAspGluPh 211
Db 594 AAACCGGACCTCCGCTTAACGAAACCCCTCCCG-----CCCCCGCGCGCGCGCGACT 538
Qy 211 eAspAlaThrValAspProArgTyrGluThrAspLeuLysAsnGluTyrIleArgHisTr 231
Db 537 CGACGCGACGCT--TCGCGCGCTTAATAAATAAATCTTTTAAGACATCCGCTAAAA 481
Qy 231 pLysGln-----ValLysLysGlyTyrCysAlaHisGlnArgArgArgGlyTh 246
Db 480 CCGAATCTTCACCCCGCACTTTACAAACAACCTTCGAGACGCGCTCGACCGCGGAC 421
Qy 246 rAlaProHisAlaArgSerIleAlaLeuIleAsnLysIleTyrGlnPro 262
Db 420 GACCCCG-----AAATCCATCGCTTAATAAATAATTTATTTATTTCCCG 378

RESULT 7
US-10-363-483A-1792
; Sequence 1792, Application US/10363483A
; Publication No. US20050064401A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Olek
; APPLICANT: Christian Piepenbrock
; APPLICANT: Kurt Berlin
; TITLE OF INVENTION: Diagnosis of illnesses or predisposition to certain
; FILE OF INVENTION: 11111111
; FILE REFERENCE: 82011
; CURRENT APPLICATION NUMBER: US/10/363,483A
; CURRENT FILING DATE: 2003-03-03
; NUMBER OF SEQ ID NOS: 40712
; SEQ ID NO 1792
; LENGTH: 1428
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; OTHER INFORMATION: CPG-Island No: 1792
US-10-363-483A-1792

Alignment Scores:
Pred. No.: 0.0412 Length: 1428
Score: 107.50 Matches: 74
Percent Similarity: 35.33% Conservative: 38
Best Local Similarity: 23.34% Mismatches: 113
Query Match: 7.35% Indels: 93
DB: 19 Gaps: 15

US-09-087-136-1 (1-275) x US-10-363-483A-1792 (1-1428)
Qy 12 LeuProGluAerGlyAspArgAsnAlaArgGlnAsnAspProLeuIleSerGlyGlyPro 31
Db 182 CTACCCGACCAAAACCCCGGAAAGCTTAAGCGCG-AAATCCGCGCTCCCTCCGCGCG 240
Qy 32 LeuPro-----LeuGluSerProSerArg----- 39
Db 241 ACCCTTAACGCGCCCGAAATTAACGACCTTAACCACTTAACCAAAACCGCGCGACATCA 300
Qy 40 -----LysLeuThrSerLeuLeuSer-TyrAspProThrValProGluSerPr 55
Db 301 AAACGCGGATCCGCGCGACCGCTCGACCTCGACATCATTAAATCC----- 352
Qy 55 oAspMetLysPheAlaArgLysArgLeuGlyAsnLeuLeuThrThrIleLysHisLysPr 75
Db 353 -----GCGCGCGCTCGACGACGAAACCTTCGCAAAACGCCCTTAACAAACCC 399
Qy 75 oSerGluIleIleGlyValLeuProGluAspTyrThrArgAlaAspGluGluProGlyAr 95
Db 400 GACC-----CCGACCCCGGACTCGACTCGACGACGACGACGACGACGACGACGAC-- 436
```

```
QY 95 gGlnGlyArgProGlyArgProArgLysMetProArgHisGlySerThrSerLe 115
Db 437 -----GACCCCTCCGACACGACCGGACGCTGAAAAAATGATGACGACGA 489
QY 115 uMetGlySerProArg-----LysThrMetThrArgAsp----- 126
Db 490 AAAAAAAAAATACGAAAAACACATCCCGAATTCAAAAAGAAACACCGACGAAAAA 549
QY 127 -----SerLysIleMetPheGlyLeuArgGlyLy 136
Db 550 AAAAAACGACGGCGCCGCGCTTTATACAAAAATCCGAACGAACTCCGCGCCG 609
QY 136 sPro-----PheGlyMetIleAlaGlyArgPheGly----- 146
Db 610 CCGACCGGACTTTCACGCAAAATCTTATTCACAAATCACTCGACGACCAATATCC 669
QY 147 -----GlnGlyTyrSerLeuGlyArgAlaTrpValLysGlyHisMetAs 161
Db 670 TCCGACGCTTCCAAACAAAACTACGCTATAATTAACTATACCTTCAAAAAAGCGAA 729
QY 161 nAsnGlyTyrGlyProIleLysAlaGlnArgThrAspTyrAlaProAsnLeuAlaValAs 181
Db 730 CCGAACGACGATCCTAAAAAATAAATACGCGCGACGCGCCGAAACCGCCCGCG-- 787
QY 181 pTyrLeuAlaCysArgGlyIleHisArgMetProArgProAspLysSerIlePro----- 199
Db 788 -----CAACCCCGCACCGGCTCCGCAAAACCGCCCGCGCGCCCGCTCC 834
QY 200 -----GlnLeuProIleValProSerArgIleAspGlyLuph 211
Db 835 AAACCCGACCTCCGCTAACGAAACCCCTCCG--CCCCCGGCGCCACGCGCGACT 891
QY 211 eAspAlaThrValAspProArgTyrGlyThrAspLeuLysAsnGlyTyrIleArgHisTr 231
Db 892 CGACCGCGAC--TCGCCCCCTATAAAAAATCTTTTAAAAACATCCCGCTAAAA 948
QY 231 pLysGln-----ValLysGlyTyrCysAlaHisGlnArgArgTrh 246
Db 949 CCGAATCTTCACCCCGACCTTTTACACAAACCTTCGACGAAACCTCGACCGCGAAC 1008
QY 246 fAlaProHisAlaArgSerIleAlaLeuIleAsnLysIleTyrGlnPro 262
Db 1009 GACCCCG-----AAATCATGCTCTAAAAAATATTTATTTCCCG 1051

RESULT 8
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
```

```
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
Alignment Scores:
Pred. No.: 3.47e+03 Length: 9025608
Score: 107.50 Matches: 62
Percent Similarity: 36.19% Conservative: 31
Best Local Similarity: 24.12% Mismatches: 93
Query Match: 7.35% Indels: 72
DB: 15 Gaps: 13

US-09-087-136-1 (1-275) x US-10-156-761-1 (1-9025608)
QY 5 AspProLeuAlaGlyPheLeuLeuProGlnAspLysArgAsnAlaArgGlnAsnAsp 24
Db 8205870 GACGATGTCAGACACCTGCTACGCCCGGAGTGCAGACCGGACCGCGGAAACGACG 8205811
QY 25 ProLeuIleSerGlyGlyPro-----LeuProLeuGlySerProSer 38
Db 8205810 ACCGAGCTCTCGGCGGTCGCGCGCGCGCCGACGACCTGCGCGCAACCGGAACCCAC 8205751
QY 39 ArgLysLeuThr-----SerLeuLeuSerTyrAspProThrValProGlySerPro 55
Db 8205750 GCACCTCCGACAGTGCACAGCCGCGCTCTCCGTCGACCAACCGACGCGCGCCG 8205691
QY 56 AspMetLysPheAlaArgLys----- 62
Db 8205690 GCGCTCGGCGGCTTCCAGCGGCGCGCGGACATGCGCGGACCGCGCGCGCGGCACTG 8205631
QY 63 ArgLeuGlyAsnLeuLeuThrThrIleLysHisProSerGlyIle-----Gly 80
Db 8205630 CGCGCAGCG-----CATCGGAGGACAGTAACCCCGCGGG 8205595
QY 81 ValLeuProGlnAspTyrThrArgAlaAspGlnLuph-----Gly 94
Db 8205594 CACCTGACACGACGACGAGTCTGTCGAGCGCTCCCGACGCGCTGCGACGAGAGT 8205535
QY 95 ArgGlnGlyArgProGlyArg-----ProArgLysMetPro-----Arg 108
Db 8205534 CGGACGCGCGCGCGCGCGCGCTGCTCGCGGCGAGATCCCGTCCCGCACGCGCG 8205475
QY 109 HisGlu-----SerSerThrSerLeuMetGlySerPr 119
Db 8205474 CACGAGAGAGACACCGCGCTACCTGTCCCGCTCGCGCACGAGCGTGCGCCACCC 8205415
QY 119 cArgLysThrMetThrArgAspSerLysIleMetPheGlnLeuArgGlyLysProPheG 139
Db 8205414 GCTCTGACAGAGCTGCGCGGCTTCACCGTGCAGCTCCAGCTCCCGACGAAACCGGCA 8205355
QY 139 uMetIleAlaGlyArgPheGlyGlyGlyTyrSerLeuGlyArgAlaTrpValLysGlyHis 159
Db 8205354 CTG-----TCGCGGTCGCGCGGACGCTCCAGCACTCCCGCACATGGGAAACGCGCG 8205301
QY 159 sMetAsnAsnGlyTyrGlyProIleLysAlaGlnArgThrAspTyrAlaProAsnLeuAl 179
Db 8205300 GTTGTCGTCGTGGGAGAGATCCCGTCTCCACACGCGCAACGAGCCGAGA----- 8205246
QY 179 aValAspTyrLeuAlaCysArgGlyIleHisArgMetProArgProAsp----- 195
Db 8205245 -----TTCGACAGCGG-CGCGAA--CATCGGCGCGCGGTCCGACGACCGCTGCA 8205197
QY 196 -----LysSerIleProGlnLeuProIleValPro 205
Db 8205196 CGGCGCGCGCGCGGACGCGGACGACCTCCGCTAGCGCGCTGCGCC 8205148

RESULT 9
US-10-200-562-190
; Sequence 190, Application US/10200562
; Publication No. US20030165819A1
; GENERAL INFORMATION:
; APPLICANT: McGowen, Patrick
; APPLICANT: Hosken, Nancy A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
```



```

; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210121.538C2
; CURRENT APPLICATION NUMBER: US/10/200.562
; CURRENT FILING DATE: 2002-07-19
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 190
; LENGTH: 9369
; TYPE: DNA
; ORGANISM: HSV2
; US-10-200-562-190

Alignment Scores:
Pred. No.: 0.539 Length: 9369
Score: 107.00 Matches: 62
Percent Similarity: 32.00% Conservative: 18
Best Local Similarity: 24.80% Mismatches: 93
Query Match: 7.32% Indels: 77
DB: 16 Gaps: 10

US-09-087-136-1 (1-275) x US-10-200-562-190 (1-9369)
Qy 13 ProgluabpGlyAspArgAsnAlaArgGlnAsnAppProleuIleSerGlyIleProleu 32
Db 8063 CCGAGGGGTGGGGTGGCGC-----CGGCACTCCCGCGCGGTGGCGCGCGGCTT 8116
Qy 33 ProleugIuserProserArgIleuTherSerleuIuserTyArgpProThValPro 52
Db 8117 CCGCGGCTCTCCGACACGCGC-----CCACCC 8146
Qy 53 GluserProabpMetIlyPheAlaArgIlyAspArgIleuTherThrilIleu 72
Db 8147 CGGAGTCCG-----CGCCCGGCTGGGTTTCTGCTTCCCTTCGCGC 8191
Qy 73 HishIserProserGluIleIleGlyValIleuProgluabpTyThrArgAlaAspGlu 91
Db 8192 CCGGCGGCTCGCGCTCGCGGGGCTTCCCGCCCGACGCGTGGCGCATCCGCGCTC 8251
Qy 92 GluPro-GlyArgGlnGlyArgProProGlyArgProArgIlyMetProArgHIsGlu 111
Db 8252 CGCCCGCGGAAGAGCGGTCTCGCCCGGGGAGACCGCGCGCGCGCGCGACAGAA 8311
Qy 111 rserTherSerleuMetGluSerProArgIlyTherMetThrArgAspSerIlyIleMetPh 131
Db 8312 CAGCGGGTCTCTCGCGACACCCCGCGGGCGCGCGCGCGCGCGCGCGCTCA 8371
Qy 131 eglIleuArgIlyAspProPheGluMetIleAlaGlyArgPheGluGluGlyTySerIle 151
Db 8372 CGCGCGCGCGCGGTGC-----CTCGCTCT 8395
Qy 151 uGlyArgAlaTrpValIlysgIlyHIsMetAsnAsnGluTyGluProIleIlyS-AlaGlnA 171
Db 8396 CGGCTCTCGTT-----GAACTCCCTCCCTCTCCCGCGGAGC 8431
Qy 171 rGThrAspTyAla-ProAsnIleuAlaValAspTyIleuAlaCyAspArgIleHIsArg 190
Db 8432 CAGCGGAGTCTCTCGCGACACCCCGCGGGCGCGCGCGCGCGCGCGCGCTCA 8491
Qy 191 Met---ProArgProAbpIlySerIleProGluIleuProIleValProserArgIleAsp 209
Db 8492 TTGCGCCCGCGACCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8524
Qy 210 GluPheAspAlaThrValAspProArgTyGluThrAspIleuIlyAsnGluTyIleArg 229
Db 8525 -----CCCGGCTT----- 8533
Qy 230 HIsTrpIlysgIlyValIlyS-----IlysgIlyTrpCyAlaHIs 241
Db 8534 ---TGGCGCCGAGCGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 8590
Qy 242 GlnArgArgArgThralaProHIs 249
Db 8591 GCCGGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8614
```

```

RESULT 10
US-10-237-551-190
; Sequence 190, Application US/10237551
; Publication No. US20030165820A1
; GENERAL INFORMATION:
; APPLICANT: Day, Craig H.
; APPLICANT: Hoeken, Nancy A.
; APPLICANT: Parsons, Joseph M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; FILE REFERENCE: 210121.538C3
; CURRENT APPLICATION NUMBER: US/10/237.551
; CURRENT FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 190
; LENGTH: 9369
; TYPE: DNA
; ORGANISM: HSV2
; US-10-237-551-190

Alignment Scores:
Pred. No.: 0.539 Length: 9369
Score: 107.00 Matches: 62
Percent Similarity: 32.00% Conservative: 18
Best Local Similarity: 24.80% Mismatches: 93
Query Match: 7.32% Indels: 77
DB: 16 Gaps: 10

US-09-087-136-1 (1-275) x US-10-237-551-190 (1-9369)
Qy 13 ProgluabpGlyAspArgAsnAlaArgGlnAsnAppProleuIleSerGlyIleProleu 32
Db 8063 CCGAGGGGTGGCGGTGGCGC-----CGGCACTCCCGCGCGGTGGCGCGCGGCTT 8116
Qy 33 ProleugIuserProserArgIleuTherSerleuIuserTyArgpProThValPro 52
Db 8117 CCGCGGCTCTCCGACACGCGC-----CCACCC 8146
Qy 53 GluserProabpMetIlyPheAlaArgIlyAspArgIleuTherThrilIleu 72
Db 8147 CGGAGTCCG-----CGCCCGGCTGGGTTTCTGCTTCCCTTCGCGC 8191
Qy 73 HishIserProserGluIleIleGlyValIleuProgluabpTyThrArgAlaAspGlu 91
Db 8192 CCGGCGGCTCGCGCTCGCGGGGCTTCCCGCCCGACGCGTGGCGCATCCGCGCTC 8251
Qy 92 GluPro-GlyArgGlnGlyArgProProGlyArgProArgIlyMetProArgHIsGlu 111
Db 8252 CGCCCGCGGAAGAGCGGTCTCGCCCGGGGAGACCGCGCGCGCGCGCGCGCGCGCTCA 8311
Qy 111 rserTherSerleuMetGluSerProArgIlyTherMetThrArgAspSerIlyIleMetPh 131
Db 8312 CAGCGGGTCTCTCGCGACACCCCGCGGGCGCGCGCGCGCGCGCGCGCGCTCA 8371
Qy 131 eglIleuArgIlyAspProPheGluMetIleAlaGlyArgPheGluGluGlyTySerIle 151
Db 8372 CGCGCGCGCGCGGTGC-----CTCGCTCT 8395
Qy 151 uGlyArgAlaTrpValIlysgIlyHIsMetAsnAsnGluTyGluProIleIlyS-AlaGlnA 171
Db 8396 CGGCTCTCGTT-----GAACTCCCTCCCTCTCCCGCGGAGC 8431
Qy 171 rGThrAspTyAla-ProAsnIleuAlaValAspTyIleuAlaCyAspArgIleHIsArg 190
Db 8432 CAGCGGAGTCTCTCGCGACACCCCGCGGGCGCGCGCGCGCGCGCGCGCGCGCTCA 8491
Qy 191 Met---ProArgProAbpIlySerIleProGluIleuProIleValProserArgIleAsp 209
Db 8492 TTGCGCCCGCGACCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8524
Qy 210 GluPheAspAlaThrValAspProArgTyGluThrAspIleuIlyAsnGluTyIleArg 229
```

```
Db      8525 -----|||
QY      230 HistTrpGlnValys-----LysGlyTrpCysAlaHis 241
Db      8534 ---TGCCCCAGCCCGGTGGCCCCGAGGAGCCTTTGTGGGTGGTGGTGGCCGAG 8590
QY      242 GlnArgArgArgThrAlaProHis 249
Db      8591 GCCCGGTGGCCGCCGCCGCCAC 8614

RESULT 11
US-10-237-551-247
; Sequence 247, Application US/10237551
; Publication No. US20030165820A1
; GENERAL INFORMATION:
; APPLICANT: Day, Craig H.
; APPLICANT: Hosken, Nancy A.
; APPLICANT: Parsons, Joseph M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; FILE REFERENCE: 210121.538C3
; CURRENT APPLICATION NUMBER: US/10/237,551
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 247
; LENGTH: 9369
; TYPE: DNA
; ORGANISM: Herpes simplex virus
US-10-237-551-247

Alignment Scores:
Pred. No.: 0.539      Length: 9369
Score: 107.00      Matches: 62
Percent Similarity: 32.00%      Conservative: 18
Best Local Similarity: 24.80%      Mismatches: 93
Query Match: 7.32%      Indels: 77
DB: 16      Gaps: 10

US-09-087-136-1 (1-275) x US-10-237-551-247 (1-9369)
QY      13 ProGlnAspGlyAspArgAsnAlaArgGlnAsnAspProLeuIleSerGlyGlyProLeu 32
Db      8063 CCGAGCGGTGGCGTCCGC-----CGGCACTCCCGCCCGGTCCGCCGCGTCCGCTT 8116
QY      33 ProLeuGlnSerProSerArgGlySerLeuThrSerLeuLeuSerTyAspProThrValPro 52
Db      8117 CCGCGCGCTCTCCGACACCGCCG-----CCACCC 8146
QY      53 GlnSerProAspMetLysPheAlaArgLysArgLeuGlyAsnLeuLeuThrThrIleLys 72
Db      8147 CGGAGTCCG-----CGCCCCCGGCTTGGGTTTCTGCTTTCGCCCTTCGCCG 8191
QY      73 HisHisProSerGlnIleIleGlyValLeuProGlnAspTyTrThrArgAlaAspGln--- 91
Db      8192 CCGGCGCGCTCGCTCGCGCGGCGCTTCCCGCCGCCACCGCTCGCGCCATCCGCGCTC 8251
QY      92 GlnPro-GlyArgGlnGlyArgProProGlyArgProArgLysMetProArgHisGlnUse 111
Db      8252 CGCCCGCGAAGAGAGCGGTCTGCTCCCGGGGAGACCGCGCGCGCGCGCCGACAA 8311
QY      111 rSerThrSerLeuMetGlnSerProArgLysThrMetThrArgAspSerLysIleMetPh 131
Db      8312 CAGCGGGTCTCTCGCGACACCGCGCGGGGCGCGCGCGCGCGCCACCGCGGCGCTCA 8371
QY      131 eGlnLeuArgGlyLysProPheGlnMetIleAlaGlyArgPheGlnGlnGlyTrpSerLe 151
Db      8372 CGCGCGCGCGCGCTCC-----CTCGCTCT 8395
QY      151 uGlyArgAlaTrpValLysGlyHisMetAsnAsnGlnLysTrpGlnProIleLys-AlaGln 171
Db      8396 CGGCTCTCT-----GAACTCTCTCTCTCTCTCCCGGAGAC 8431
```

```
QY      171 rGlnAspTyAla-ProAsnLeuAlaValAspTyLeuAlaCysArgGlnIleHisArg 190
Db      8432 CAGCGACACAGCGCGCGGTTCGCCCGGACGCGGCGAGTCCGCTTCTCCGGGCC 8491
QY      191 Met---ProArgProAspLysSerIleProGlnLeuProIleValProSerArgIleAsp 209
Db      8492 TTGCCCCCGGACCTCCGCGCGTCCAGACCAACC----- 8524
QY      210 GlnPheAspAlaThrValAspProArgTyGlnThrAspLeuLysAsnGlnLysTrpIleArg 229
Db      8525 -----CCCCGCTT----- 8533
QY      220 HistTrpGlnValys-----LysGlyTrpCysAlaHis 241
Db      8534 ---TGCCCCAGCCCGGTGGCCCCGAGGAGCCTTTGTGGGTGGTGGTGGCCGAG 8590
QY      242 GlnArgArgArgThrAlaProHis 249
Db      8591 GCCCGGTGGCCGCCGCCGCCAC 8614

RESULT 12
US-09-827-688-8/c
; Sequence 8, Application US/09827688
; Publication No. US20030165476A1
; GENERAL INFORMATION:
; APPLICANT: KINSEY, FRANK
; APPLICANT: BHOAGAL, BALBIR
; TITLE OF INVENTION: MACROAGREGATED PROTEIN CONJUGATES AS ORAL GENETIC IMMUNIZATION D
; FILE REFERENCE: P01949US/1004014
; CURRENT APPLICATION NUMBER: US/09/827,688
; PRIOR APPLICATION NUMBER: 60/195,680
; PRIORITY FILING DATE: 2001-04-06
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 154746
; TYPE: DNA
; ORGANISM: HERPESVIRUS 2
US-09-827-688-8

Alignment Scores:
Pred. No.: 20.4      Length: 154746
Score: 107.00      Matches: 62
Percent Similarity: 32.00%      Conservative: 18
Best Local Similarity: 24.80%      Mismatches: 93
Query Match: 7.32%      Indels: 77
DB: 10      Gaps: 10

US-09-087-136-1 (1-275) x US-09-827-688-8 (1-154746)
QY      13 ProGlnAspGlyAspArgAsnAlaArgGlnAsnAspProLeuIleSerGlyGlyProLeu 32
Db      72872 CCGAGGGGTGGCTCCGC-----CGGCACTCCCGCCCGGTCCGCCGCGCTT 72819
QY      33 ProLeuGlnSerProSerArgLysLeuThrSerLeuLeuSerTyAspProThrValPro 52
Db      72818 CCGCGGCTCTCCGACACCGCCG-----CCACCC 72789
QY      53 GlnSerProAspMetLysPheAlaArgLysArgLeuGlyAsnLeuLeuThrThrIleLys 72
Db      72788 CGGAGTCCG-----CGCCCCCGGCTTGGGTTTCTGCTTTCGCCCTTCGCCG 72744
QY      73 HisHisProSerGlnIleIleGlyValLeuProGlnAspTyTrThrArgAlaAspGln--- 91
Db      72743 CCGGCGCGCTCGCTCGCGGGGCGCTTCCCGCCCGACGCTCGCGGCCATCCGCGCTC 72684
QY      92 GlnPro-GlyArgGlnGlyArgProProGlyArgProArgLysMetProArgHisGlnUse 111
Db      72683 CGCGCGCGAAGAGCGGTCTCTCCCGGGGAGACCGCGCGCGCGCGCCGACAA 72624
```

[illegible]

```

; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.: 5.17e+03 Length: 9025608
Score: 106.00 Matches: 75
Percent Similarity: 32.48% Conservative: 27
Best Local Similarity: 23.89% Mismatches: 95
Query Match: 7,25% Indels: 117
DB: 15 Gaps: 14

US-09-087-136-1 (1-275) x US-10-156-761-1 (1-9025608)
QY 24 AspproLeuIIseRgIyGlyProleuProleuGluSerProserArglyLeuThrSer 43
DB 2493789 GACCAATCATCGCGCGACGCTCAAGCTCTGACGCGCATCGACCGGAGTCC----- 2493842
QY 44 LeuLeuSerTyRAspproThrValProGluSerProAspMetLeuPheAlaArglySer 63
DB 2493843 -----GACTTCAACCAACCGGAGGCGCAAGACTAC-----CGG 2493875
QY 64 LeuGlyAsnLeuLeuThrThrIleIlyshIshIProserGluIleIleGlyValLeuPro 83
DB 2493876 TTGGGTGACCGGCTCGCCACCATCATGTGTCGCCCGCGGAGC-----TGGCAGCTGACG 2493929
QY 84 GluAspTyRThrArgAlaAspGluGluPro----- 93
DB 2493930 GAGGAACACCTGAGATTGACGCGCGCGGCGGTAACCGGCTCTCTGCACTTCGAGCTG 2493989
QY 94 -----GlyArgGln----- 96
DB 2493990 TACTTCTTCCACTGCGCGGAGCGGACGATCGACGCGGACAGACCGTACTTCACTG 2494049
QY 96 ----- 96
DB 2494050 CCGAAGCTGAGAACAGATACGAAGCCGCTGTGGAACGAGTCTTCCTCCGCCGAG 2494109
QY 97 -----GlyArgProProGlyArgProArglySerMetProArgHisGluSerSerThr 113
DB 2494110 GAGCTGCTCGGATCCCCCGGAGCCATCGGCGGACCGCTGCTCATCGACGATCAACC 2494169
QY 114 SerLeuMetGluSerProArglySerThrMetThrArgAspSerLeuIleMetPheGluLeu 133
DB 2494170 GCCGCCGCGGAGATG-----GAGGAGATCTCTTACGAGCTG 2494205
QY 134 ArgGlyLysProPheGluMetIleAlaGlyArgPheGluGluGlyTyrSerIleuGlyArg 153
DB 2494206 CGCGAACAACAGCGCGGCTCAACCGGAGCGCTGGAATATCTTTCAGCTTGATCAAG 2494265
QY 154 AlaTrpValLysGlyHisMetAsnGluTyrGluProIleIlyAlaGlnArgThrAsp 173
DB 2494266 ACCCTT-----GGGCGACCGGACCGCATCTTCTGCTCGCGGACCGGTCAGCATG 2494319
QY 174 TyrAlaProAsnLeu---AlaValAspTyRLeuAlaCysArgGluIleHis---Arg-Me 191
```

```

DB 2494320 ACCGACCCCTTCTCGCGCGCTACACCGAACTGCTGTGCGCACTTCCACCGCGCGGC 2494379
QY 191 tPrArgProAspLysSerIlePProGluLeuProIleValProSerArgIleAspGluPh 211
DB 2494380 GCCCAGCCCATCGCGGCGCATGTCGCGC-----AGTG 2494412
QY 211 eAspAlaThrValAspProArgTyRgIuThrAspLeuLysAsnGluTyrIleArgHisTr 231
DB 2494413 CCGCGCAGGATTCGCGCGCGCGGAGG-----CGGCCCTG 2494448
QY 231 pLysGlnValLysLysGlyTyrCysAlaHisGlnArgArgThrAlaProHisAlaAr 251
DB 2494449 GCGA-----AGTCCCGCTGGAACAGAGCGCGGAGCGGCTTTCAGCGCT-- 2494500
QY 251 gSerIleAlaLeuLeuAsnLysIleTyRgIuProGly----- 263
DB 2494501 -----CTGGGTGGCGCCACCCCGGCTGTGCGC 2494529
QY 264 -----GluSerLysThrValGluGlnAlaLeuGly 273
DB 2494530 ATGTCCGCGGAGGTCTTCAACGTGTCTCGCGGAGCGGC 2494569

RESULT 15
US-10-156-761-5319/c
; Sequence 5319, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMDA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5319
; LENGTH: 1041
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1041)
US-10-156-761-5319

Alignment Scores:
Pred. No.: 0.0532 Length: 1041
Score: 105.00 Matches: 52
Percent Similarity: 32.54% Conservative: 30
Best Local Similarity: 20.63% Mismatches: 84
Query Match: 7,18% Indels: 87
DB: 15 Gaps: 11

US-09-087-136-1 (1-275) x US-10-156-761-5319 (1-1041)
QY 17 AsppArgAsnAlaArgGlnAsnAsp-----ProleuIleSerGlyGlyProleuProleu 34
DB 552 GATATCATCCCGCCGACGATCCGACGCGCGGACCGGACCGGAGACCTCCCGCCG 493
QY 35 GluSerProSerArgLysLeuThrSerLeuLeuSerTyRAspProThrValProGluSer 54
DB 492 CGGCACCTTGC----- 481
QY 55 ProAspMetLysPheAlaArgLysArgLeuGlyAsnLeuLeuThrThrIleLysHisHis 74
```

```
Db 480 -----CTGGCCACCAC 469
Qy 75 ProSerGluIleIleGlyValLeuProGluAspTyrThrArgAlaAspGluGluProGly 94
Db 468 -----CGCCCGGAA 460
Qy 95 ArgGlnGlyArgProGlyArgProArgLysMetPro-----ArgHis 109
   |||::||| |||::||| |||::|||
Db 459 CGCCGTGGCAACCGGACGCCGACGCCGCCGCCGCCGCCGAAACGACAGCTGTCGCCGAT 400
Qy 110 GluSerSerThrSerLeuMetGluSerProArgLysThrMetThrArgAspSerLysIle 129
   ::::|::| |::|::| |::|::|
Db 399 CGCGACCCAGTTCACTGCGGTGCTCCAGCGCCAGTACTGCGGAGAGAGCCCAAGT 340
Qy 130 MetHeGluLeuArgGlyLysProHeGlu-----MetIleAlaGlyArgPheGluGlu 147
   |||::||| |||::||| |||::|||
Db 339 CGTGAATCCCGGCTGGGCTGCTGCGACACTGCTGATGCGCCGCCGCGCACGCGGC 280
Qy 148 GluTyrSerLeuGlyArgAlaTrpValLysGlyHisMetAsnAsnGluTyr---GluPro 166
   |||::||| |||::||| |||::|||
Db 279 GCAGCTTCGCAAGC-----GCAGACGAACGGTACGCTGACCCG 241
Qy 167 IleLys-----AlaGlnArgThrAspTyrAlaProAsnLeu 178
   ::::|::| |::|::| |::|::|
Db 240 GTGCGCGGCGCCACCCGCTCACCCGGGTGCGCCACCGCTCGACCAAGCCGCGAGTT 181
Qy 179 AlaValAspTyrLeuAlaCysArgGluIleHisArgMetProArgProAspLysSerIle 198
   ::::|::| |::|::| |::|::|
Db 180 GTGCCCGCGACATCGGACAGGTGATGTCGGGTGTCGCCGCCACCCA-----CCCGTG 127
Qy 199 ProGluLeuProIleValProSerArgIleAspGluPheAspAlaThrValAspProArg 218
   |||::||| |||::||| |||::|||
Db 126 CCAAGTCGCTCCGGCACAACCCGGTGCCTC-----GACCCGCAC 88
Qy 219 TyrGluThrAspLeuLysAsnGluTyrIleArgHisTrpLysGlnValLysGlyTrp 238
   ::::|::| |::|::| |::|::|
Db 87 CACGACCCCGTGTCTCGAGGGCTCCGATCCGCCACTCCCGCACT-CCGCGC----- 35
Qy 239 CysAlaHisGlnArgArgThrAlaProHisAla 250
   |||::||| |||::||| |||::|||
Db 34 --GCTCCCGAACCGCTCGAACACACACCGCACGCA 2
```

Search completed: April 9, 2005, 12:56:00
Job time : 3925 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_p1ue_p2n model

Run on: April 9, 2005, 08:02:07 ; Search time 195 Seconds

(without alignments)
2307.570 Million cell updates/sec

Title: US-09-087-136-1
Perfect score: 1462
Sequence: 1 MSEIDPLAEFLPEDGDGNA.....INKIYQGESKTVQALGLI 275

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 81818359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+g2n.model -DEV=xjh
-Q=/cgn2_1/USPRO.spool/US09087136/runat_07042005_163616_932/app_query.fasta_1.455
-DB=Issued_Patents_NA -QFMT=fastrap -SUFFIX=rn1 -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MTN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09087136@cgn2_1.1177@runat_07042005_163616_932 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCtus.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	109.5	7.5	1707	1 US-08-790-309-1	Sequence 1, Appli
C 2	109.5	7.5	1707	3 US-09-250-585A-1	Sequence 190, App
3	107	7.3	9369	4 US-10-237-551-190	Sequence 14633, A
4	107	7.3	9369	4 US-10-237-551-247	Sequence 14636, A
5	107	7.3	57280	4 US-09-949-016-11796	Sequence 14637, A
6	107	7.3	57280	4 US-09-949-016-12843	Sequence 14638, A
7	107	7.3	57280	4 US-09-949-016-12844	Sequence 14639, A
8	107	7.3	57280	4 US-09-949-016-12845	Sequence 14640, A
9	107	7.3	57280	4 US-09-949-016-13542	Sequence 14641, A
10	107	7.3	57280	4 US-09-949-016-13543	Sequence 14642, A
11	107	7.3	57280	4 US-09-949-016-13544	Sequence 14643, A
12	107	7.3	57280	4 US-09-949-016-13545	Sequence 14644, A

13	107	7.3	57280	4 US-09-949-016-14633	Sequence 14633, A
14	107	7.3	57280	4 US-09-949-016-14634	Sequence 14634, A
15	107	7.3	57280	4 US-09-949-016-14635	Sequence 14635, A
16	107	7.3	57280	4 US-09-949-016-14636	Sequence 14636, A
17	107	7.3	57280	4 US-09-949-016-14637	Sequence 14637, A
18	107	7.3	57280	4 US-09-949-016-14638	Sequence 14638, A
19	107	7.3	57280	4 US-09-949-016-14639	Sequence 14639, A
20	107	7.3	57280	4 US-09-949-016-14640	Sequence 14640, A
C 21	107	7.3	154746	4 US-09-827-688-8	Sequence 8, Appli
C 22	106.5	7.3	1908	4 US-09-252-991A-16529	Sequence 16529, A
C 23	106.5	7.3	2247	4 US-09-252-991A-1856	Sequence 1856, Ap
C 24	104.5	7.1	2172	4 US-09-252-991A-1708	Sequence 1708, Ap
C 25	104.5	7.1	2241	4 US-09-252-991A-2016	Sequence 2016, Ap
C 26	104.5	7.1	2888	3 US-08-765-907A-1	Sequence 1, Appli
C 27	104.5	7.1	2888	4 US-09-987-614A-1	Sequence 1, Appli
C 28	104.5	7.1	4039	4 US-09-902-540-696	Sequence 696, App
C 29	104	7.1	3289	4 US-09-587-945-1	Sequence 1, Appli
C 30	103.5	7.1	5894	3 US-08-665-259-24	Sequence 24, Appli
C 31	103.5	7.1	5894	3 US-08-762-500-24	Sequence 24, Appli
C 32	103.5	7.1	6447	4 US-09-949-016-4925	Sequence 4925, Ap
C 33	103.5	7.1	6525	3 US-08-762-500-74	Sequence 74, Appli
C 34	98	6.7	1710	2 US-08-912-794-1	Sequence 1, Appli
C 35	98	6.7	2676	4 US-09-252-991A-14053	Sequence 14053, A
C 36	98	6.7	2874	4 US-09-252-991A-14101	Sequence 14101, A
C 37	97.5	6.7	1404	4 US-09-252-991A-12486	Sequence 12486, A
C 38	97.5	6.7	71989	3 US-09-443-501A-2	Sequence 2, Appli
C 39	97	6.6	2178	4 US-09-252-991A-4641	Sequence 4641, Ap
C 40	97	6.6	2721	4 US-09-252-991A-4237	Sequence 4237, Ap
C 41	96	6.6	68750	3 US-09-335-409-1	Sequence 1, Appli
C 42	96	6.6	68750	3 US-09-568-102-1	Sequence 1, Appli
C 43	96	6.6	68750	3 US-09-567-969-1	Sequence 1, Appli
C 44	96	6.6	68750	3 US-09-568-480-1	Sequence 1, Appli
C 45	96	6.6	68750	3 US-09-568-486-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-790-309-1/c
Sequence 1, Application US/08790309
Patent No. 5786174
GENERAL INFORMATION:
APPLICANT: Demirjian, David C
APPLICANT: Casadaban, Malcolm J
APPLICANT: Pagratzis, Nikos
APPLICANT: Vonstein, Veronika
TITLE OF INVENTION: Thermophile Gene Transfer
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Allegretti & Witcoff, Ltd.
STREET: 10 S. Wacker Drive Suite 3000
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/790,309
FILING DATE: 28-JAN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/265,522
FILING DATE: 24 JUN 1994
ATTORNEY/AGENT INFORMATION:
NAME: McDonnell, John J
REGISTRATION NUMBER: 26,949
REFERENCE/DOCKET NUMBER: 93,1000
TELECOMMUNICATION INFORMATION:

```
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1707 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 262..810
US-08-790-309-1

Alignment Scores:
Pred. No.: 0.0356 Length: 1707
Score: 109.50 Matches: 81
Percent Similarity: 36.92% Conservative: 39
Best Local Similarity: 24.92% Mismatches: 94
Query Match: 7.49% Indels: 113
DB: 1 Gaps: 19

US-09-087-136-1 (1-275) x US-08-790-309-1 (1-1707)

QY 2 SerGluileaspPro-----LeuAgluPhe-----Leu 11
DB 1137 GCGGAAGTAGACCCGCTGCGCTCCCGCTCCAGAGACACACACCTCCCGACCTC 1078
QY 12 LeuProGluaspGlyAspArgasnAlaArgGlnasnAspProLeu----- 26
DB 1077 CTCGCCGAGGCGCGC-----CGGGAAGAGACCCGGTTCCTCCCTCCACCTCGGG 1030
QY 27 -----LleSerGlyGlyProLeu----- 32
DB 1029 GCTTTCAGGTGACGATAAGCCATCGGGGGCGGGGGGAAAGAGCGAAAGCTCGAGG 970
QY 33 -----ProLeuGluSerProSerArg---LysLeuThrSer 43
DB 969 AGCCTCGGGGCTCTCCGAGAGGGGCGCTTAAGAGACTCTCGAGGCTGGGAGAGCTCC 910
QY 44 LeuLeuSerTyrrAspProThrVal----- 51
DB 909 CCCCTGAGCCACCGGGGACGAGTCTCCCTCGCGGCTCTCCGCGGCGCGAGGCGCAG 850
QY 52 ProGluSerProAspMetLysPheAlaArgLysArgGlnGlyAsnLeuLeuThrThrIle 71
DB 849 GAGAAAGCGCGAGAAAGGCGAGGAGGCGACCTTAACCTCTC----- 799
QY 72 LysHisHisPro-----SerGluileileGlyValLeuProGluAspTyx 86
DB 798 CAAGGGCACCCCTCCGCGGACAGGGGCGAGGCTCTCGGATATCTG----- 751
QY 87 ThrArgAlaaspGluGluProGlyArgGlnGlyArgProProGlyArgProArgLysMet 106
DB 750 GGGAACTTCGAAGGAGAGGCGCGGAGGCGACGCGGAAGGCGCGCTCGCCCGCT 691
QY 107 ProArgHisGluSerSerThrSerLeuMetGluSerProArgLysThrMetThrArgAsp 126
DB 690 CCGGTCACAGATGGCCCCACGCC---CACCAAAACCCGCGCGGCTCGCGCGCGCG 634
QY 127 SerLysIleMetPheGluLeuArgGly----- 135
DB 633 GATGCGCTTGGGAGCGTCTCCCGGTGTGTTACACGTCCTCCACCGCGCAAGAGCGTCG 574
QY 136 -----LysProPheGluMetIleAlaGlyArgPheGluGluGlySer 150
DB 573 CCCGGGTTACAGGTAGCCCTTGGCGATGAGCATCCCTCCCTT---CCGTCCTCTCG 517
QY 151 LeuGlyArgAlaTrpValIleGlyHisMetAsnGlnGlyArgProIleLysAlaGln 170
DB 516 GCGAAGAGGCGCGGCGCGAGGCTTGGCCACCGAAGGAAAG----- 472
QY 171 ArgThrAspTyrrAlaProAsnLeuAlaValAspTyrrLeuAlaCysArgGluileHisArg 190
```

```
DB 471 -----GACCAAGCCCCGATGCGCGGCGCAT-----GACGAAGTCCAC--- 433
QY 191 MetProArgProAspLysSerIleProGluLeuProIleValProSerArgIleAspGlu 210
DB 432 CTCTCT-GGTCTCAAGAGCTTTCGCAAGAGCTCCCGACGCGCTCGGCGT-----AAA 380
QY 211 PheAspAlaThrValaAspProArgTyrrGluThrAspLeuLysAsnGlyTyrrIleArgHis 230
DB 379 GGGGATGCTGG-----AGGAGGGCGCGCCGACCTGCATAA----- 348
QY 231 TrpLysGlnValIleLysLysGlyTrpCysAlaHisIleGlnArgArgThrAlaProHisAla 250
DB 347 -----AAGAAAGGGGAGTGCATCCCGAGGCGAGAGGAAGTGGCTCTAGA 300
QY 251 ArgSerIleAlaLeu 255
DB 299 AGAGCCCCCGTCTC 285

RESULT 2
US-09-250-585A-1/c
Sequence 1, Application US/09250585A
Patent No. 6350591
GENERAL INFORMATION:
APPLICANT: Weber, Mark J.
APPLICANT: Demirjian, David C.
APPLICANT: Casadaban, Malcolm J.
APPLICANT: Pagratlis, Nikos
APPLICANT: Vonstein, Veronika
TITLE OF INVENTION: Theophilile Gene Transfer
FILE REFERENCE: 93,1000-D
CURRENT APPLICATION NUMBER: US/09/250,585A
CURRENT FILING DATE: 1999-11-16
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 1707
TYPE: DNA
ORGANISM: Thermus flavus
FEATURE:
NAME/KEY: CDS
LOCATION: (262)..(810)
US-09-250-585A-1

Alignment Scores:
Pred. No.: 0.0356 Length: 1707
Score: 109.50 Matches: 81
Percent Similarity: 36.92% Conservative: 39
Best Local Similarity: 24.92% Mismatches: 94
Query Match: 7.49% Indels: 113
DB: 3 Gaps: 19

US-09-087-136-1 (1-275) x US-09-250-585A-1 (1-1707)

QY 2 SerGluileaspPro-----LeuAgluPhe-----Leu 11
DB 1137 GCGGAAGTAGACCCGCTGCGCTCCCGCTCCAGAGACACACACCTCCCGACCTC 1078
QY 12 LeuProGluaspGlyAspArgasnAlaArgGlnasnAspProLeu----- 26
DB 1077 CTCGCCGAGGCGCGC-----CGGGAAGAGACCCGGTTCCTCCCTCCACCTCGGG 1030
QY 27 -----LleSerGlyGlyProLeu----- 32
DB 1029 GCTTTCAGGTGACGATAAGCCATCGGGGGCGGGGGGAAAGAGCGAAAGCTCGAGG 970
QY 33 -----ProLeuGluSerProSerArg---LysLeuThrSer 43
DB 969 AGCCTCGGGGCTCTCCGAGAGGGGCGCTTAAGAGACTCTCTGAGGCTCGGAGAGACTCC 910
QY 44 LeuLeuSerTyrrAspProThrVal----- 51
DB 909 CCCCTGAGCCACCGGGGACGAGTCTCCCTCGCGGCTCTCCGAGGCGCGAGGCGCAG 850
```



```

Qy 52 ProGluSerProAspMetLysPheAlaArgLysArgLeuGlyLysLeuLeuThrThrIle 71
Db 849 GAGAAAGAGCGAGAAAGAGGAGCGAGAGGAGCGACCTTAACCTCTC----- 799
Qy 72 LysHisPro-----SerGluIleIleGlyValLeuProGluAspTyr 86
Db 798 CAAGGGGACCCCTCCCGGAGAGGGGAGGAGCGCTCCGAGATATCTG----- 751
Qy 87 ThrArgAlaSerGluGluProGlyArgGlnGlyArgProProGlyArgProArgLysMet 106
Db 750 GGGAACTCCCAAGGCGAGAGGCGCGAGAGGCGAGCGGAGCGCGCTGCCCGCCCT 691
Qy 107 ProArgHisGluSerSerThrSerLeuMetGluSerProArgLysThrMetThrArgAsp 126
Db 690 CCGGTCACGATGGCCCCACAGCC--CACAAAGACCCCGCGGCGCTCCGCGCGCG 634
Qy 127 SerLysIleMetPheGluLeuArgGly----- 135
Db 633 GATGCGCTTGGGAGCGCTCTCCCGGTGTACACGCTCCACCGCGCAAGAGCGCTCG 574
Qy 136 -----LysProPheGluMetIleAlaGlyArgPheGluGluGlySer 150
Db 573 CCGGAGTTACAGGTGAGCCCTTGGCGATGAGCATCCCTCCCTT---CCGTCTCTCG 517
Qy 151 LeuGlyArgAlaTyrValLysGlyHisMetAsnAsnGlyTyrGluProIleLysAlaGln 170
Db 516 GCGAAGAGGCGCGCGCGCGCGCTTGCACCAAGAGGAGG----- 472
Qy 171 ArgThrAspTyrAlaProAsnLeuAlaValAspTyrLeuAlaCysArgGluIleHisArg 190
Db 471 -----GACCAAGCGCGCGATGCGCGGCGAT-----GACGAAGTCCAC--- 433
Qy 191 MetProArgProAspLysSerIleProGluLeuProIleValProSerArgIleAspGlu 210
Db 432 CTCTCT-CGTCTCAAGAGCTTCCCAAGGCTCCCGACGCGCTCGCGG---AAA 380
Qy 211 PheAspAlaThrValAspProArgTyrGluThrAspLeuLysAsnGlyTyrIleArgHis 230
Db 379 GGGGATGCTGG-----AGAGAGGCGCGCGCGCGCAAA----- 348
Qy 231 TrpLysGlnValLysLysGlyTyrCysAlaHisGlnArgArgThrAlaProHisAla 250
Db 347 -----AAGAAAGGAGAGTGCATCCCGACGCGAGAGAGAGAGTGCCTCTAGA 300
Qy 251 ArgSerIleAlaLeu 255
Db 299 AGAGCGCGCGCTCTC 285

RESULT 3
US-10-237-551-190
; Sequence 190, Application US/10237551
; Patent No. 6821519
; GENERAL INFORMATION:
; APPLICANT: Day, Craig H.
; APPLICANT: Hoeken, Nancy A.
; APPLICANT: Parsons, Joseph M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; FILE REFERENCE: 210121.538C3
; CURRENT APPLICATION NUMBER: US/10/237,551
; CURRENT FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 190
; LENGTH: 9369
; TYPE: DNA
; ORGANISM: HSV2
US-10-237-551-190

Alignment Scores:
Pred. No.: 0.863 Length: 9369
Score: 107.00 Matches: 62

```

```

Percent Similarity: 32.00% Conservative: 18
Best Local Similarity: 24.80% Mismatches: 93
Query Match: 7.32% Gaps: 77
DB: 4 Gaps: 10

US-09-087-136-1 (1-275) x US-10-237-551-190 (1-9369)

Qy 13 ProGluAspGlyAspArgAsnAlaArgGlnAsnAspProLeuIleSerGlyLysProLeu 32
Db 8063 CCGAGGAGTGGCGTCCG-----CGGACTCCCGCGCGCGCTGCGCGCGCGCTT 8116
Qy 33 ProLeuGluSerProSerArgLysLeuThrSerLeuLeuSerTyrAspProThrValPro 52
Db 8117 CCGGCGCTCTCCCGACAGCGCG----- 8146
Qy 53 GluSerProAspMetLysPheAlaArgLysArgLeuGlyLysLeuLeuThrThrIleLys 72
Db 8147 CGGAGTCCG-----CGCGCGCGCTTGGGTTCGCTTGGCGCGCTCCGC 8191
Qy 73 HisHisProSerGluIleIleGlyValLeuProGluAspTyrThrArgAlaAspGlu--- 91
Db 8192 CCGGCGCGCTCGGCTCGCGGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCTC 8251
Qy 92 GluPro-GlyArgGlnGlyArgProProGlyArgProArgLysMetProArgHisGluSe 111
Db 8252 CGCGCGCGAGAGAGCGGCTGCTCCCGGCGCGAGCCCGCGCGCGCGCGCGCGCGCA 8311
Qy 111 rSerThrSerLeuMetGluSerProArgLysThrMetThrArgAspSerLysIleMetPh 131
Db 8312 CAGCGGCTCCCTCGCGCACACCCCGCGGCGCGCGCGCGCGCGCGCGCGCGCTCA 8371
Qy 131 eGluLeuArgGlyLysProPheGluMetIleAlaGlyArgPheGluGluGlyTyrSerIle 151
Db 8372 CGCGCGCGCGGCTGC-----CTCGCTCT 8395
Qy 151 uGlyArgAlaTyrValLysGlyHisMetAsnAsnGlyTyrGluProIleLys-AlaGln 171
Db 8396 CGGCTCTGTT-----GAACTCCCTCCCTCTCCCGCGGAGC 8431
Qy 171 rGThrAspTyrAla-ProAsnLeuAlaValAspTyrLeuAlaCysArgGluIleHisArg 190
Db 8432 CAGCGAGCAAGCGCGCGGTTCCCGCGAGCGGCGGCGAGTCCGCTTCCGCGGCGC 8491
Qy 191 Met---ProArgProAspLysSerIleProGluLeuProIleValProSerArgIleAsp 209
Db 8492 TTGCGCGCGCGACTCCGCGCTCCAGACAGCC----- 8524
Qy 210 GluPheAspAlaThrValAspProArgTyrGluThrAspLeuLysAsnGlyTyrIleArg 229
Db 8525 -----CCCGGCTT----- 8533
Qy 230 HisTrpLysGlnValLys-----LysGlyTyrCysAlaHis 241
Db 8534 ---TGCGCGAGCGCGCGGTGCGCGAGAGCGAGCTTGTGGGTGGGTGCGCGGAG 8590
Qy 242 GlnArgArgArgThrAlaProHis 249
Db 8591 GCCCGTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8614

RESULT 4
US-10-237-551-247
; Sequence 247, Application US/10237551
; Patent No. 6821519
; GENERAL INFORMATION:
; APPLICANT: Day, Craig H.
; APPLICANT: Hoeken, Nancy A.
; APPLICANT: Parsons, Joseph M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; FILE REFERENCE: 210121.538C3
; CURRENT APPLICATION NUMBER: US/10/237,551
; CURRENT FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 254

```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 247
; LENGTH: 9369
; TYPE: DNA
; ORGANISM: Herpes simplex virus
US-10-237-551-247
```

Alignment Scores:

Pred. No.:	0.863	Length:	9369
Score:	107.00	Matches:	62
Percent Similarity:	32.00%	Conservative:	18
Best Local Similarity:	24.80%	Mismatches:	93
Query Match:	7.32%	Indels:	77
DB:	4	Gaps:	10

US-09-087-136-1 (1-275) x US-10-237-551-247 (1-9369)

```
QY 13 ProGluAAspGlyAAspArgAsnAlaArgGlnAsnAspProLeuIleSerGlyGlyProLeu 32
DB 8063 CCGAGGGGTGGCGTCCG-----CGGCACTCCCGCGCGTCCGCGCGCGCGTCCGCTT 8116
QY 33 ProLeuGluSerProSerArgLysLeuThrSerLeuLeuSerTyrAspProThrValPro 52
DB 8117 CCGCGCTCTCCCGACCAAGCCG-----CCACC 8146
QY 53 GluSerProAspMetLysPheAlaArgLysGlyAsnLeuLeuThrThrIleLys 72
DB 8147 CGAGAGTCG-----CGCCCGCGCTTGGTTCGCTTGGCCCTCCGC 8191
QY 73 HisHisProSerGluIleIleGlyValLeuProGluAspTyrThrArgAlaAspGlu--- 91
DB 8192 CCGCGCCCGCTCGCGCGCGCGCTTCCCGCCCGACCGCTCGCGCGCCATCCGCGCTC 8251
QY 92 GluPro-GlyArgGlnIlyArgProProGlyArgProArgLysMetProArgHisGlySe 111
DB 8252 CGCCCGCGGAGAGAGCGGTGTCTCCCGGGGCGACCGCGCGCGCGCGCGCGCGCGCGCTCA 8371
QY 111 rSerThrSerLeuMetGluSerProArgLysThrMetThrArgAspSerLysIleMetPh 131
DB 8312 CAGCGGGTCTCTCGCGACACCCCGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCA 8371
QY 131 eGluLeuArgGlyLysProPheGluMetIleAlaGlyArgPheGluGluGlyTyrSerLe 151
DB 8372 CGCGCGCGCGCGCGCG-----CTCGCTCT 8395
QY 151 uGlyArgAlaTyrValLysGlyHisMetAsnGlnTyrGluProIleLys-AlaGlnA 171
DB 8396 CGGCTCTCGT-----GAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8431
QY 171 rGThrAspTyrAla-ProAsnLeuAlaValAspTyrLeuAlaCysArgGluIleHisArg 190
DB 8432 CAGCGGACCGACCGCGCGCGGTTCGCGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 8491
QY 191 Met---ProArgProAspLysSerIleProGluLeuProIleValProSerArgLysAsp 209
DB 8492 TTGCCCCCGGACCTCCGCGCTCCAGACACCC-----8524
QY 210 GluPheAspAlaThrValAspProArgTyrGluThrAspLeuLysAsnGlnTyrIleArg 229
DB 8525 -----CCCCGCTT-----8533
QY 230 HisTrpLysGlnValLys-----LysGlyTyrCysValHis 241
DB 8534 ---TGGCCCCAGCGCGGTTCGCGCGACGAGCTTGTGCGGTGGGTGATGCCGAG 8590
QY 242 GlnArgArgArgThrAlaProHis 249
DB 8591 GCCCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8614
RESULT 5
US-09-949-016-11796
; Sequence 11796 Application US/09949016
; Patent No. 6812339
```

```
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11796
; LENGTH: 57280
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-11796
```

Alignment Scores:

Pred. No.:	13	Length:	57280
Score:	107.00	Matches:	56
Percent Similarity:	35.34%	Conservative:	26
Best Local Similarity:	24.14%	Mismatches:	80
Query Match:	7.32%	Indels:	70
DB:	4	Gaps:	13

US-09-087-136-1 (1-275) x US-09-949-016-11796 (1-57280)

```
QY 11 LeuLeuProGluAspGlyAAspArgAsnAlaArgGlnAsnAspProLeuIleSerGlyGly 30
DB 1368 CTACTACCCAGAGTCTGAAAAGGCTATTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCC 1427
QY 31 ProLeuProGluGluSerProSerArgLysLeuThrSerLeuLeuSerTyrAspProThr 50
DB 1428 CCGCTCTCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1469
QY 51 ValProGluSerProAspMetLysPheAlaArgLysArgGlyAsnLeuLeuThrThr 70
DB 1470 GCGCTCTCGCTCTGAG---AATGCGCAGCAAGCGCTCGCGGAGCGTGGAGCG 1526
QY 71 IleLysHisHisProSerGluIleIleGlyValLeuProGluAspTyrThrArgAlaAsp 90
DB 1527 -----GGCGCGCTCTCT-----CTGAGTCCGCGCGGTCCGACG-----1559
QY 91 GluGluProGlyArgGlnIlyArgProProGlyArgProArgLysMetProArgHisGln 110
DB 1560 -----CCGAGCGAGTCCGCGACGCTGACATCCGCGTACATCCCTGCGC-----1604
QY 111 SerSerThrSerLeuMetGluSerProArgLysThrMetThrArgAspSerLysIleMet 130
DB 1605 -----TCTCTGTGACGCGA-----1619
QY 131 PheGluLeuArgGlyLysProPheGluMetIleAlaGlyArgPheGluGluGlyTyrSer 150
DB 1620 -----AAGGCGAAAGAGCAGGACCTGCAAGCGCGC-----GGCAGCGCG 1658
QY 151 LeuGlyArgAla-----TrpValLysGlyHisMet 160
DB 1659 GTAGAGAGAGGCTCTCGTACGTGCGGACCGCGGCTCTCTGAGGT-----1706
QY 161 AsnAsnGlnTyrGluProIleLysAlaGlnArgThrAspTyrAlaProAsnLeuAlaVal 180
DB 1707 -----CATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1742
QY 181 AspTyrLeuAlaCysArgGlnIleHis---ArgMetProArgProAspLysSerIlePro 199
DB 1743 -----TCTGCGCGCGCGCTTTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1793
QY 200 GluLeuProIleValProSerArgTyrIleAspGluPheAspAlaThrValAspProArgTyr 219
DB 200 -----1793
```

Db 1794 CAAGCCCAAGCCGCGGCTCCGCACTGACTCTTACAGCGCGCGCC 1853
Qy 220 GlnthrAspLeuLyAsnGluTyrIleArgHisTrp 231
Db 1854 GCGCGACCCCGCGAGCCAG---AGCAAGACTG 1886

RESULT 6
US-09-949-016-12843
; Sequence 12843, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12843
; LENGTH: 57280
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12843

Alignment Scores:
Pred. No.: 13 Length: 57280
Score: 107.00 Matches: 56
Percent Similarity: 35.34% Conservative: 26
Best Local Similarity: 24.14% Mismatches: 80
Query Match: 7.32% Indels: 70
DB: 4 Gaps: 13

US-09-087-136-1 (1-275) x US-09-949-016-12843 (1-57280)

Qy 11 LeuLeuProGluAaspGlyAspArgAsnAlaArgGlnAsnAaspProLeuIleSerGly 30
Db 1368 CTACTACCCCGAGATGCTGAAAAGGCTATTCTGCTCTTCCCTGCTGACTCTTCC 1427
Qy 31 ProLeuProLeuGluSerProSerArgLySerLeuSerLeuSerTyrAaspProThr 50
Db 1428 CCGCTCTGCTGCTCCGCGCCAGCAGA-----GGCGGCTCTCCACT 1469
Qy 51 ValProGluSerProAaspMetLyPheAlaArgLyAspGlyAsnLeuLeuThr 70
Db 1470 GCGGCTCTGCTGCTGAG---AAATGGCGCAGCAAGCGCTCCGCGAGAGCTGAGCG 1526
Qy 71 IleLyshHisProSerGluIleIleGlyValLeuProGluAaspTyrThrArgAlaAsp 90
Db 1527 -----GGCGGCTCTTCT-----CTGAGTCCGCGGCGGTCCGACC----- 1559
Qy 91 GluGluProGlyArgGlnGlyArgProProGlyArgProArgLyMetProArgHisGlu 110
Db 1560 -----CCGAGCCGATCGCGCAGACTGCAATCCGCGAGCATCCCTGCC----- 1604
Qy 111 SerSerThrSerLeuMetGluSerProArgLySerThrMetThrArgAaspSerLyIleMet 130
Db 1605 -----TCTCTGTGACGCGGA----- 1619
Qy 131 PheGluLeuArgGlyArgProPheGluMetIleAlaGlyArgPheGluGluGlySer 150
Db 1620 -----AAGCGCAAAAGCAGGACTGCAAGCGGC-----GCCACCGG 1658
Qy 151 LeuGlyArgAla-----TPPVAllyGlyHisMet 160
Db 1659 GTAGAAAGAGCGGCTCTGCTAGTGCAGCGAGCCGCGCTCTCGGTT----- 1706

Qy 161 AsnAsnGluTyrGluProIleLyAlaGlnArgThrAaspTyrAlaProAsnLeuAlaVal 180
Db 1707 -----CATCCCCCGCGCACCTCGGGGTCCGACCCGCG----- 1742
Qy 181 AspTyrLeuAlaCyArgGluIleHis---ArgMetProArgProAaspLySerIlePro 199
Db 1743 -----TCTCTCGGCGCTTTTGGCGGCACCCGCGCTCCGACCCGCTCTCC 1793
Qy 200 GluLeuProIleValProSerArgLyIleAaspGluPheAspAlaThrValAaspProArgTyr 219
Db 1794 CAAGCCCAAGCCGCGGCGGCTCCGCACTGACTCTTACAGCGCGCGCGCC 1853
Qy 220 GlnthrAspLeuLyAsnGluTyrIleArgHisTrp 231
Db 1854 GCGCGACCCCGCGAGCCAG---AGCAAGACTG 1886

RESULT 7
US-09-949-016-12844
; Sequence 12844, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12844
; LENGTH: 57280
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12844

Alignment Scores:
Pred. No.: 13 Length: 57280
Score: 107.00 Matches: 56
Percent Similarity: 35.34% Conservative: 26
Best Local Similarity: 24.14% Mismatches: 80
Query Match: 7.32% Indels: 70
DB: 4 Gaps: 13

US-09-087-136-1 (1-275) x US-09-949-016-12844 (1-57280)

Qy 11 LeuLeuProGluAaspGlyAspArgAsnAlaArgGlnAsnAaspProLeuIleSerGly 30
Db 1368 CTACTACCCCGAGATGCTGAAAAGGCTATTCTGCTCTTCCCTGCTGACTCTTCC 1427
Qy 31 ProLeuProLeuGluSerProSerArgLySerLeuSerLeuSerTyrAaspProThr 50
Db 1428 CCGCTCTGCTGCTCCGCGCCAGCAGA-----GGCGGCTCTCCACT 1469
Qy 51 ValProGluSerProAaspMetLyPheAlaArgLyAspGlyAsnLeuLeuThr 70
Db 1470 GCGGCTCTGCTGCTGAG---AAATGGCGCAGCAAGCGCTCCGCGAGAGCTGAGCG 1526
Qy 71 IleLyshHisProSerGluIleIleGlyValLeuProGluAaspTyrThrArgAlaAsp 90
Db 1527 -----GGCGGCTCTTCT-----CTGAGTCCGCGGCGGTCCGACC----- 1559
Qy 91 GluGluProGlyArgGlnGlyArgProProGlyArgProArgLyMetProArgHisGlu 110
Db 1560 -----CCGAGCCGATCGCGCAGACTGCAATCCGCGTACATCCCTGCC----- 1604
Qy 111 SerSerThrSerLeuMetGluSerProArgLySerThrMetThrArgAaspSerLyIleMet 130

Db 1605 -----TCTCTGTGACGGCA----- 1619
Qy 131 PhegluLeuarglyLysProPhegluMetIleaglYargPhegluLugluTyrSer 150
Db 1620 -----AAGGCAAAAGGACGAGGACTGCAAGCGGGC-----GCGCACCGG 1658
Qy 151 LeuGLYargAla-----TyrValLysglYHisMet 160
Db 1659 GTAGGAAGAGCGGCTCTGGCTAGAGTGGCGGACCGCGGCTCTGGGTT----- 1706
Qy 161 AsnAsnGluTyrGluProIleLysAlaGlnargThrAspTyrAlaProAsnLeuAlaVal 180
Db 1707 -----CCATCCCGCGCGCACCTCGGGGTCCGACCCGGC----- 1742
Qy 181 AspTyrLeuAlaCysargGluIleHis---ArgMetProargProAspLysSerIlePro 199
Db 1743 -----TCTGCGGGGCGCTTTTGGCGCGACCGCTCCCGACCGGCTCTCTCC 1793
Qy 200 GluLeuProIleValProSerargIleAspGluPheAspAlaThrValAspProargTyr 219
Db 1794 CAAGCCCAACCGGCGCCAGCGCGGTCCCGACCTGACTGACTCTACAGCGCGCGCC 1853
Qy 220 GluThrAspLeuLysAsnGluTyrIleargHisTyr 231
Db 1854 GCGCGCACCGCGCGACGCGAG---AGCAAGCACTGG 1886
RESULT 8
US-09-949-016-12846
Sequence 12846, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C1001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12846
LENGTH: 57280
TYPE: DNA
ORGANISM: Human
US-09-949-016-12846
Alignment Scores:
Pred. No.: 13 Length: 57280
Score: 107.00 Matches: 56
Percent Similarity: 35.34% Conservative: 26
Best Local Similarity: 24.14% Mismatches: 80
Query Match: 7.32% Indels: 70
DB: 4 Gaps: 13
US-09-087-136-1 (1-275) x US-09-949-016-12846 (1-57280)
Qy 11 LeuLeuProGluAspGlyAspArgAsnAlaArgGlnAsnAspProLeuIleSerGlyGly 30
Db 1368 CTACTACCCAGAGATGCTGAAGAAAGGCTATTCTGCTCTCTCCCTGCTGACGCTTCC 1427
Qy 31 ProLeuProLeuGluSerProSerargLysIleuthrSerLeuLeuSerTyrAspProThr 50
Db 1428 CCGGCTCTGCTCGCGCGCGCGAGCA-----GGCGCGGCTCCCACT 1469
Qy 51 ValProGluSerProAspMetLysPheAlaArgLysargLeuGlyAsnLeuLeuThrThr 70
Db 1470 GCGGCTCTGCTGCTGAG---AAATGGCGCAGCAAGCGCTCCGCGGAGAGCGCTGAGCGG 1526

Qy 71 IleLysHisSerProSerGluIleIleGlyValLeuProGluAspTyrThrArgAlaAsp 90
Db 1527 -----GGGCGCCCTTCT-----CTGATCCGCGGGGTCCACCC----- 1559
Qy 91 GluGluProGlyArgGlnGlyArgProProGlyLysProargLysMetProargHisGlu 110
Db 1560 -----CCGACCAGGTGGCGCCAGACCTGCATCCCGCTGATATCCCTGCCC----- 1604
Qy 111 SerSerThrSerLeuMetGluSerProargLysThrMetThrArgAspSerLysIleMet 130
Db 1605 -----TCTCTGTGACGGCA----- 1619
Qy 131 PhegluLeuArgglyLysProPhegluMetIleaglYargPhegluLugluTyrSer 150
Db 1620 -----AAGGCAAAAGGACGAGGACTGCAAGCGGGC-----GCGCACCGG 1658
Qy 151 LeuGLYargAla-----TyrValLysglYHisMet 160
Db 1659 GTAGGAAGAGCGGCTCTGGCTAGAGTGGCGGACCGCGGCTCTGGGTT----- 1706
Qy 161 AsnAsnGluTyrGluProIleLysAlaGlnargThrAspTyrAlaProAsnLeuAlaVal 180
Db 1707 -----CCATCCCGCGCGCACCTCGGGGTCCGACCCGGC----- 1742
Qy 181 AspTyrLeuAlaCysargGluIleHis---ArgMetProargProAspLysSerIlePro 199
Db 1743 -----TCTGCGGGGCGCTTTTGGCGCGACCGCTCCCGACCGGCTCTCTCC 1793
Qy 200 GluLeuProIleValProSerargIleAspGluPheAspAlaThrValAspProargTyr 219
Db 1794 CAAGCCCAACCGGCGCCAGCGCGGTCCCGACCTGACTGACTCTACAGCGCGCGCC 1853
Qy 220 GluThrAspLeuLysAsnGluTyrIleargHisTyr 231
Db 1854 GCGCGCACCGCGCGACGCGAG---AGCAAGCACTGG 1886
RESULT 9
US-09-949-016-13542
Sequence 13542, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: C1001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13542
LENGTH: 57280
TYPE: DNA
ORGANISM: Human
US-09-949-016-13542
Alignment Scores:
Pred. No.: 13 Length: 57280
Score: 107.00 Matches: 56
Percent Similarity: 35.34% Conservative: 26
Best Local Similarity: 24.14% Mismatches: 80
Query Match: 7.32% Indels: 70
DB: 4 Gaps: 13
US-09-087-136-1 (1-275) x US-09-949-016-13542 (1-57280)
Qy 11 LeuLeuProGluAspGlyAspArgAsnAlaArgGlnAsnAspProLeuIleSerGlyGly 30
Db 1368 CTACTACCCAGAGATGCTGAAGAAAGGCTATTCTGCTCTCTCCCTGCTGACGCTTCC 1427
Qy 31 ProLeuProLeuGluSerProSerargLysIleuthrSerLeuLeuSerTyrAspProThr 50
Db 1428 CCGGCTCTGCTCGCGCGCGCGAGCA-----GGCGCGGCTCCCACT 1469
Qy 51 ValProGluSerProAspMetLysPheAlaArgLysargLeuGlyAsnLeuLeuThrThr 70
Db 1470 GCGGCTCTGCTGCTGAG---AAATGGCGCAGCAAGCGCTCCGCGGAGAGCGCTGAGCGG 1526

```
Dh 1368 CTACTACCCGAGATGTGAAAAGGCTATTCTGCTCTTCCCTGCTGACGTCTTCC 1427
Qy 31 ProleupProleugluserProserArglyseuthrSerleuLeuserTyAspProthr 50
Dh 1428 CCGCTCTGCTCGCGCCGACGAGA-----GGCCGGCGCTTCCACT 1469
Qy 51 ValProgluserProAspMetLysPheAlaArglysaArgleuglyAsnleuLeuthrThr 70
Dh 1470 GCGCTCTGCTGCTGAG---AAATGGCGCAGCAAGCCCTCCGCGAAGCTGAGGCG 1526
Qy 71 IlelyshshsPserSerGluIleIleIleValleuProgluAspTyThrArgAlaAsp 90
Dh 1527 -----GGCGCCCTTCT-----CTGAGTCCGCGGCGGCGCACC----- 1559
Qy 91 GluGlupProglArgGlnGlyArgProProglArgProArglyMetProArgHisGlu 110
Dh 1560 -----CCGACCCAGTCCGCGCAGACTGATCCCGCGAGCATCCCTGCC----- 1604
Qy 111 SerSerThrSerLeuMetGluSerProArglyThrMetThrArgAspSerlyIleMet 130
Dh 1605 -----TCTCTGTGACGCGA----- 1619
Qy 131 PheGluLeuArgGlyArgProPheGluMetIleAlaGlyArgPheGluGluGlySer 150
Dh 1620 -----AAGGCGAAAGGAGGAGCTGCAGAGCGGC-----GCCACCGG 1658
Qy 151 LeuGlyArgAla-----TTPVallyGlyHisMet 160
Dh 1659 GTAGGAAAGCGGCTCTGCTGAGTGCAGGACCGCGGCTCTGCTGTT----- 1706
Qy 161 AsnAsnGluTyrgluProIlelysaGlnArgThrAspTyraIaProAsnleuAlaVal 180
Dh 1707 -----CCATCCCCCGCGCGCACCTCGGGGTCCGACCCGCG----- 1742
Qy 181 AspTyrlleuAlaCyArgGluIleHis---ArgMetProArgProAspLysSerIlePro 199
Dh 1743 -----TCCTGCGCGGCTTTTCGCGCGCACCCGCTCCGACCCGCTCTCTCC 1793
Qy 200 GluLeuProIleValProSerArgIleAspGluPheAspAlaThrValAspProArgTy 219
Dh 1794 CAAGCCCAAGCGCGCCCAAGCGCGCTCCGCACTGAGCTCTACACGCGCGCGCC 1853
Qy 220 GluThrAspLeuLysAsnGluTyrlleArgHisTrp 231
Dh 1854 GGCGCGACCCCGCGCAGCAG---AGCAAGCACTGG 1886

RESULT 10
US-09-949-016-13543
; Sequence 13543, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: PasteSeq for Windows Version 4.0
; SEQ ID NO 13543
; LENGTH: 57280
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13543

Alignment Scores:
Pred. No.: 13 Length: 57280
```

```
Score: 107.00 Matches: 56
Percent Similarity: 35.34% Conservative: 26
Best Local Similarity: 24.14% Mismatches: 80
Query Match: 7.32% Indels: 70
DB: 4 Gaps: 13

US-09-087-136-1 (1-275) x US-09-949-016-13543 (1-57280)
Qy 11 LeuLeuProgluAspArglyAspArgAsnAlaArgGlnAsnAspProleuIleSerGlyGly 30
Dh 1368 CTACTACCCGAGATGTGAAAAGGCTATTCTGCTCTTCCCTGCTGACGTCTTCC 1427
Qy 31 ProleupProleugluserProserArglyseuthrSerleuLeuserTyAspProthr 50
Dh 1428 CCGCTCTGCTCGCGCCGACGAGA-----GGCCGGCGCTTCCACT 1469
Qy 51 ValProgluserProAspMetLysPheAlaArglysaArgleuglyAsnleuLeuthrThr 70
Dh 1470 GCGCTCTGCTGCTGAG---AAATGGCGCAGCAAGCGCTCCGCGAAGCTGAGGCG 1526
Qy 71 IlelyshshsPserSerGluIleIleIleValleuProgluAspTyThrArgAlaAsp 90
Dh 1527 -----GGCGCCCTTCT-----CTGAGTCCGCGGCGGCGCACC----- 1559
Qy 91 GluGlupProglArgGlnGlyArgProProglArgProArglyMetProArgHisGlu 110
Dh 1605 -----CCGACCCAGTCCGCGCAGACTGATCCCGCGAGCATCCCTGCC----- 1604
Qy 111 SerSerThrSerLeuMetGluSerProArglyThrMetThrArgAspSerlyIleMet 130
Dh 1605 -----TCTCTGTGACGCGA----- 1619
Qy 131 PheGluLeuArgGlyArgProPheGluMetIleAlaGlyArgPheGluGluGlySer 150
Dh 1620 -----AAGGCGAAAGGAGGAGCTGCAGAGCGGC-----GCCACCGG 1658
Qy 151 LeuGlyArgAla-----TTPVallyGlyHisMet 160
Dh 1659 GTAGGAAAGCGGCTCTGCTGAGTGCAGGACCGCGGCTCTGCTGTT----- 1706
Qy 161 AsnAsnGluTyrgluProIlelysaGlnArgThrAspTyraIaProAsnleuAlaVal 180
Dh 1707 -----CCATCCCCCGCGCGCACCTCGGGGTCCGACCCGCG----- 1742
Qy 181 AspTyrlleuAlaCyArgGluIleHis---ArgMetProArgProAspLysSerIlePro 199
Dh 1743 -----TCCTGCGCGGCTTTTCGCGCGCACCCGCTCCGACCCGCTCTCTCC 1793
Qy 200 GluLeuProIleValProSerArgIleAspGluPheAspAlaThrValAspProArgTy 219
Dh 1794 CAAGCCCAAGCGCGCCCAAGCGCGCTCCGCACTGAGCTCTACACGCGCGCGCC 1853
Qy 220 GluThrAspLeuLysAsnGluTyrlleArgHisTrp 231
Dh 1854 GGCGCGACCCCGCGCAGCAG---AGCAAGCACTGG 1886

RESULT 11
US-09-949-016-13544
; Sequence 13544, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
```

```

; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13544
; LENGTH: 57280
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13544

Alignment Scores:
Pred. No.: 13          Length: 57280
Score: 107.00         Matches: 56
Percent Similarity: 35.34%      Conservative: 26
Best Local Similarity: 24.14%   Mismatches: 80
Query Match: 7.32%             Indels: 70
DB: 4                      Gaps: 13

US-09-087-136-1 (1-275) x US-09-949-016-13544 (1-57280)

QY      11 LeuLeuProGluAspGlyAspArgAsnAlaArgGlnAsnAspProLeuIleSerGlyGly 30
DB      1368 CTACTACCCAGAGATGCTGAAAAGGCTATTCTGCTCTCTCTCCCTGCTCAGCTCTTCC 1427
QY      31 ProLeuProLeuGluSerProSerArgGlySerLeuThrSerLeuLeuSerTyraAspProThr 50
DB      1428 CCGCTCTGCTGCTCGCGCCCAAGAGA-----GGCCGCGCTCCACT 1469
QY      51 ValProGluSerProAspMetIysPheAlaArgIysArgLeuGlyAsnLeuLeuThrThr 70
DB      1470 GCGCTCTGCTGCTGAG--AAATGGCGCAGCAAGCGCTCCGCGAGAGCTGAGAGCG 1526
QY      71 IleYsHisHisProSerGluIleIleGlyValLeuProGluAspTyraArgAlaAsp 90
DB      1527 -----GGGCGCCCTTCT-----CTAGTCCGCGGGGTCCGACC----- 1559
QY      91 GluGluProGlyArgGlnGlyArgProGlyArgProGlyArgProGlySerMetProArgHisGlu 110
DB      1560 -----CGAGCCAGTGGCCAGACCTGATCCCGGAGTACATCCCTGCCC----- 1604
QY      111 SerSerThrSerLeuMetGluSerProArgGlySerMetThrArgAspSerIysIleMet 130
DB      1605 -----TCTCTGTGACGCGA-----TCTCTGTGACGCGA----- 1619
QY      131 PheGluLeuArgGlyIysProPheGluMetIleAlaGlyArgPheGluGluGluIlyrSer 150
DB      1620 -----AAGGGCAAAAGGACGAGCATCGAAGCGGC-----GCGCACCGG 1658
QY      151 LeuGlyArgAla-----TTPValIysGlyHisMet 160
DB      1659 GTAGAGAGAGCGCTGCTGCGTAGTGGCGGACCGCGGCTCTGGGT----- 1706
QY      161 AsnAsnGluTyrgluProIleYsAlaGlnArgThrAspTyraIleProAsnLeuAlaVal 180
DB      1707 -----CCATCCCGCGCGCAGCTCGGGGTCCGACCGCGC----- 1742
QY      181 AspTyraLeuAlaCysArgGluIleHis--ArgMetProArgProAspIysSerIlePro 199
DB      1743 -----TCTGCGGGGCTTTTCGCGCGACCCGCTCCGACCGCGACCGCGCTCTCC 1793
QY      200 GluLeuProIleValProSerArgGlyLeuAspGlnPheAspAlaThrValAspProArgTy 219
DB      1794 CAAGCCCAAGCGCGCAAGCGCGCTCCGCGACCTGAGCTCTTACACGCGCGCGCGCC 1853
QY      220 GluThrAspLeuYsAsnGluTyraIleArgHisTyr 231
DB      1854 GCGCGACCGCGCGAGCAG--AGCAAGACATCG 1886

RESULT 12
US-09-949-016-13545
; Sequence 13545, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13545
; LENGTH: 57280
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13545

Alignment Scores:
Pred. No.: 13          Length: 57280
Score: 107.00         Matches: 56
Percent Similarity: 35.34%      Conservative: 26
Best Local Similarity: 24.14%   Mismatches: 80
Query Match: 7.32%             Indels: 70
DB: 4                      Gaps: 13

US-09-087-136-1 (1-275) x US-09-949-016-13545 (1-57280)

QY      11 LeuLeuProGluAspGlyAspArgAsnAlaArgGlnAsnAspProLeuIleSerGlyGly 30
DB      1368 CTACTACCCAGAGATGCTGAAAAGGCTATTCTGCTCTCTCTCCCTGCTCAGCTCTTCC 1427
QY      31 ProLeuProLeuGluSerProSerArgGlySerLeuThrSerLeuLeuSerTyraAspProThr 50
DB      1428 CCGCTCTGCTGCTCGCGCCCAAGAGA-----GGCCGCGCTCCACT 1469
QY      51 ValProGluSerProAspMetIysPheAlaArgIysArgLeuGlyAsnLeuLeuThrThr 70
DB      1470 GCGCTCTGCTGCTGAG--AAATGGCGCAGCAAGCGCTCCGCGAGAGCTGAGAGCG 1526
QY      71 IleYsHisHisProSerGluIleIleGlyValLeuProGluAspTyraArgAlaAsp 90
DB      1527 -----GGGCGCCCTTCT-----CTAGTCCGCGGGGTCCGACC----- 1559
QY      91 GluGluProGlyArgGlnGlyArgProGlyArgProGlyArgProGlySerMetProArgHisGlu 110
DB      1560 -----CGAGCCAGTGGCCAGACCTGATCCCGGAGTACATCCCTGCCC----- 1604
QY      111 SerSerThrSerLeuMetGluSerProArgGlySerMetThrArgAspSerIysIleMet 130
DB      1605 -----TCTCTGTGACGCGA-----TCTCTGTGACGCGA----- 1619
QY      131 PheGluLeuArgGlyIysProPheGluMetIleAlaGlyArgPheGluGluGluIlyrSer 150
DB      1620 -----AAGGGCAAAAGGACGAGCATCGAAGCGGC-----GCGCACCGG 1658
QY      151 LeuGlyArgAla-----TTPValIysGlyHisMet 160
DB      1659 GTAGAGAGAGCGCTGCTGCGTAGTGGCGGACCGCGGCTCTGGGT----- 1706
QY      161 AsnAsnGluTyrgluProIleYsAlaGlnArgThrAspTyraIleProAsnLeuAlaVal 180
DB      1707 -----CCATCCCGCGCGCAGCTCGGGGTCCGACCGCGC----- 1742
QY      181 AspTyraLeuAlaCysArgGluIleHis--ArgMetProArgProAspIysSerIlePro 199
DB      1743 -----TCTGCGGGGCTTTTCGCGCGACCCGCTCCGACCGCGACCGCGCTCTCC 1793
QY      200 GluLeuProIleValProSerArgGlyLeuAspGlnPheAspAlaThrValAspProArgTy 219
DB      1794 CAAGCCCAAGCGCGCAAGCGCGCTCCGCGACCTGAGCTCTTACACGCGCGCGCGCC 1853
QY      220 GluThrAspLeuYsAsnGluTyraIleArgHisTyr 231
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: April 9, 2005, 07:56:56 ; Search time 583 Seconds
(without alignments)
2792.331 Million cell updates/sec

Title: US-09-087-136-1
Perfect score: 1462
Sequence: 1 MSEIDPLAEFLPLPEDGDGNA.....INKIYQGESKTVQALGLI 275

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_model -DEV=xlh
-Q=/cgn2_1/USPTO.epool/US09087136/runat_07042005_163615_895/app_query.fasta_1.455
-DB=N Geneseq_16Dec04 -QEMT=fastcap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTPMT=pro -NORM=ext -HEAPSIZE=500 -MITLEN=0 -MAXLEN=200000000
-USER=US09087136 @CGN_1_1_644@runat_07042005_163615_895 -NCPD=6 -ICPU=3
-NO_MMAP -LARGEOUDERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database : N_Geneseq_16Dec04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1453	99.4	954	2	AAV72859
2	109.5	7.5	1707	6	AAV32780
3	109.5	7.5	1707	6	ABA92233
4	109	7.5	1315	3	AAC46404
5	109	7.5	1318	3	AAC39522

C	6	107.5	7.4	1428	6	ABQ15200	Abq15200 Oligonuc
C	7	107.5	7.4	1428	6	ABQ15201	Abq15201 Oligonuc
C	8	107.5	7.4	6906	6	ABN80010	Abn80010 Human che
C	9	107	7.3	9369	10	ADG75175	Adg75175 Human her
C	10	107	7.3	9369	10	ADG75118	Adg75118 Human her
C	11	107	7.3	117213	2	AAV62176	Aav62176 HSV-2 str
C	12	107	7.3	154746	6	AAD25519	Aad25519 Human her
C	13	106.5	7.3	1908	11	ABD17925	Abd17925 Pseudomon
C	14	106.5	7.3	2247	11	ABD03252	Abd03252 Pseudomon
C	15	105	7.2	2262	11	ACN44556	Acn44556 Pseudomon
C	16	104.5	7.1	2172	11	ABD03104	Abd03104 Pseudomon
C	17	104.5	7.1	2241	11	ABD03412	Abd03412 Pseudomon
C	18	104.5	7.1	2888	2	AAT59268	Aat59268 Streptomy
C	19	104.5	7.1	29870	8	AAD36874	Aad36874 Streptomy
C	20	104	7.1	3289	4	AAH26570	Aah26570 Human win
C	21	104	7.1	6458	6	ABK48984	Abk48984 Genomic D
C	22	103.5	7.1	5894	2	AAT85474	Aat85474 HABC3 CDN
C	23	103.5	7.1	5894	2	AAV16307	Aav16307 CDNA enc
C	24	103.5	7.1	6445	12	ADQ85442	Adq85442 Human tum
C	25	103.5	7.1	6470	4	AAH57445	Aah57445 Human tum
C	26	103.5	7.1	6491	3	AAZ94761	Aaz94761 Human ATP
C	27	103.5	7.1	6491	3	ABL66808	Abi66808 Lung canc
C	28	103.5	7.1	6491	6	ABK84439	Abk84439 Human CDN
C	29	103.5	7.1	6491	12	ADL61160	Adl61160 Human ATP
C	30	103.5	7.1	6491	13	ADR24698	Adr24698 Breast ca
C	31	103.5	7.1	6525	2	AAV16345	Aav16345 CDNA enc
C	32	103.5	7.1	6608	9	ACH03810	Ach03810 Human CDN
C	33	102.5	7.0	1474	4	AAH59796	Aah59796 Ptopionib
C	34	102.5	7.0	1474	8	ACF64725	Acf64725 Ptopionib
C	35	101.5	6.9	939	6	ABQ40537	Abq40537 Oligonuc
C	36	101.5	6.9	2775	13	ADT43632	Adt43632 Bacterial
C	37	101.5	6.9	110000	6	ABA90521	Abp90521 of
C	38	101.5	6.9	1201	10	ADC87218	Adc87218 Human GPC
C	39	100.5	6.9	2577	8	ADA70383	Ada70383 Rice gene
C	40	100.5	6.9	2209	4	AAH10704	Aah10704 Human CDN
C	41	100	6.8	2209	13	ADQ87393	Adq87393 Human tum
C	42	100	6.8	2566	4	AAH17667	Aah17667 Human CDN
C	43	100	6.8	9994	4	AAH85191	Aac85191 S. avermi
C	44	100	6.8	984	10	ABZ69821	Abz69821 Beta tran
C	45	99.5	6.8				

ALIGNMENTS

RESULT 1	
AAV72859	AAV72859 standard; CDNA; 954 BP.
XX	XX
XX	AAV72859;
AC	AC
XX	XX
DT	29-MAR-1999 (first entry)
XX	XX
DE	Caenorhabditis elegans synMuv gene lin-37 CDNA.
XX	XX
KW	LIN-37; synthetic multivulvar; SynMuv; signal transduction; animal model;
KW	tumour suppressor; retinoblastoma; cancer; cancer; cell proliferation;
KW	gene therapy; ss.
XX	XX
OS	Caenorhabditis elegans.
XX	XX
FH	Key Location/Qualifiers
FT	CDS 1..828
FT	/*tag= a
XX	XX
XX	XX
PD	03-DEC-1998.
XX	XX
PF	28-MAY-1998; 98MO-US011043.
XX	XX
PR	28-MAY-1997; 97US-0047996P.
XX	XX
PA	(MASI) MASSACHUSETTS INST TECHNOLOGY.

XX Horvitz HR, Ceol C, Lu X;
 PI WPI; 1999-045362/04.
 DR P-PSDB; AAW83386.
 XX
 PT Novel LIN-37, -35, -55, -52, -53 and -54 gene from *C. elegans* - useful
 PT for treating diseases associated with altered levels of cell
 PT proliferation, e.g. carcinomas.
 XX
 PS Claim 10; Fig 3; 70pp; English.
 XX
 CC This cDNA sequence codes for LIN-37 (see AAW83386), a novel 32 kDa
 CC hydrophilic protein of *Caenorhabditis elegans*. lin-37 cDNA was isolated
 CC from the Okkema embryonic cDNA library. The lin-37 transcript is about 1
 CC kb in size and is present both in embryonic and mixed-stage RNAs. The lin
 CC -35 gene is a novel synthetic multivulvar (symbuv) gene involved in cell
 CC fate and cell proliferation, and is part of a pathway that may be used as
 CC a genetic and biochemical model system for tumour suppression and cancer
 CC in mammals. Symbuv pathway genes and proteins may be used to identify
 CC genes which are part of the mammalian pathway and to identify genes,
 CC proteins and therapeutic compounds which modulate this pathway. Pure
 CC nucleic acids (see AAW72859-65) encoding *C. elegans* Symbuv polypeptides
 CC selected from LIN-37, -35, -55, -52, -53, -54 and B2F-1 (see AAW83386-
 CC 92), are new. Also claimed are: (1) nucleic acids (see AAW72866-67)
 CC encoding the mouse and human homologues of *C. elegans* LIN-54; (2) vectors
 CC containing the nucleic acids; (3) transgenic cells; (4) a pure mammalian
 CC Symbuv polypeptide, and (5) an antibody which binds to a Symbuv family
 CC protein. The Symbuv nucleic acids and polypeptides can be used to
 CC diagnose and treat, especially by gene therapy, conditions involving
 CC altered levels of cell proliferation, e.g. Symbuv-associated carcinomas
 XX
 SQ Sequence 954 BP; 310 A; 208 C; 183 G; 253 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 7,26e-124 Length: 954
 Score: 1453.00 Matches: 274
 Percent Similarity: 99.64% Conservative: 0
 Best Local Similarity: 99.64% Mismatches: 1
 Query Match: 99.38% Indels: 0
 DB: 2 Gaps: 0

US-09-087-136-1 (1-275) x AAW72859 (1-954)

QY 1 MetSerGluIleAspProLeuAlaGluPheLeuProGluAspGlyAspArgAsnAla 20
 DB 1 ATGTGAGAAATGATCCACTTCCGAGTCTTCTCCAGAAAGCGAGATCGAAATGCT 60
 QY 21 ArgGlnAsnAspProLeuIleSerGlyGlyProLeuProLeuGluSerProSerArgLys 40
 DB 61 CGTCAAAATGATCCATGATTAAGCGAGGTCACATTCATTGGAATGCCAAGCAGAAA 120
 QY 41 LeuThrSerLeuLeuSerTyrAspProThrValProGluSerProAspMetLysPheAla 60
 DB 121 CTCACATCCCTGTTATCTTATGATCCGACAGTCCGGAGTCACCGGATATATAATTCGCC 180
 QY 61 ArgLysArgLeuGlyAsnLeuLeuThrThrIleLysHisIleProSerGluIleIleGly 80
 DB 181 AGAAAAAGCTGTGGGAAATCTGCTGCACCAATAAACATCCATCGGAAATATATGGA 240
 QY 81 ValLeuProGluAspTyrThrArgAlaAspGluGluProGlyArgGlnGlyArgProPro 100
 DB 241 GTACCTCCGAGAAATTAATCTGCTGATGAAGAGCCCGGCGCCCAAGGACGTCCACCA 300
 QY 101 GlyArgProArgLysMetProArgHisGluSerSerThrSerLeuMetGluSerProArg 120
 DB 301 GTGTGCTCTCTTAAGATCCCGGTACCAATCTTCACTTATGGAATCCACACGC 360
 QY 121 LysThrMetThrArgAspSerLysIleWeePheGluLeuArgGlyLysProPheGluMet 140
 DB 361 AAGACTATGACTCGTGAATTTAAATATATGTTGAATTTGGGTGAAAAACATTCGAAATG 420
 QY 141 IleIleGlyArgPheGluGluGluTyrSerLeuGlyArgAlaTrpValLysGlyHisMet 160

DB 421 ATAGCTGACCTTTTGAAGAAATATTCCTTGTGAGCATGGTTAAAGGACATG 480
 QY AsnAsngluTyrGluProIleLysAlaGlnArgThrAspTyrAlaProAsnLeuAlaVal 180
 DB 481 AATAATGAATATGAACCAATAAAGCTCAAAAGGACAGACTATGCAACCAATCTGGCTGT 540
 QY 181 AspTyrLeuAlaCysArgGluIleHisArgMetProArgProAspLysSerIleProGlu 200
 DB 541 GATTATCTTGATGCTCGGAGATTCATGAAATGCCAGCTCCAGTAATCAATCATCTCGAG 600
 QY 201 LeuProIleValProSerArgIleAspGluPheAspAlaThrValAspProArgTyrGlu 220
 DB 601 CTGCCAATGTTCCATCATGATGATGAATTCGACGCTACAGTCGATCCAAAGATATGAA 660
 QY 221 ThrAspLeuLysAsnGluTyrIleArgHisTyrPheGlnValLysLysGlyTyrCysAla 240
 DB 661 ACAGATTTGAAAAATGAATATCATTCGATTTGGAACAACTCAAAAAGCTTGCTGTCT 720
 QY 241 HisGlnArgArgGlnThrAlaProHisAlaArgSerIleAlaLeuIleAsnLysIleTyr 260
 DB 721 CATCAAGTCGTGCGACTGCTCCCATGCAAGAGCATATGACATTAATCAACAAATCTAC 780
 QY 261 GlnProGluGluSerLysThrValGluGlnAlaLeuGlyLeuIle 275
 DB 781 CAGCTGGAAGTCTGAATACTGTGAGCAAGCACTTGCTTAT 825

RESULT 2
 AAW32780/C
 ID AAW32780 standard; DNA; 1707 BP.
 XX
 AC AAW32780;
 XX
 DT 17-OCT-2003 (revised)
 DT 07-OCT-1998 (first entry)
 XX
 DE Thermus flavus pyre gene.
 XX
 KW Thermus flavus pyre gene; kanamycin resistance; leucine gene;
 KW heat tolerance; thermophile; ss.
 XX
 OS Thermus thermophilus.
 XX
 FH Key Location/Qualifiers
 FT CDS 262..810
 FT /tag= a
 FT /product= "pyre protein"

XX US5786174-A.
 XX 28-JUL-1998.
 XX 28-JAN-1997; 97US-00790309.
 XX 24-JUN-1994; 94US-00265522.
 XX (THER-) THERMOGEN INC.
 PI Bigratis NC, Weber JM, Vonstein V, Demirjian DC, Casadaban MJ;
 XX WPI; 1998-436521/37.
 DR P-PSDB; AAW49024.
 XX
 PT Transformation of *Thermus* species with exogenous DNA - allows
 PT thermophilic expression of genes not normally associated with high
 PT temperatures.
 XX
 PS Disclosure; Fig 2; 24pp; English.

CC The present sequence represents the *Thermus flavus* pyre gene. The
 CC invention provides a method for thermophile gene transfer which involves
 CC producing recombinant thermophilic host cells which contain exogenous DNA
 CC under control of *Thermus* chromosomes, especially kanamycin resistance,

CC under control of *Thermus thermophilus* leucine and pyre genes. The
 CC integration of desired genes into specific locations allows the exogenous
 CC protein to become thermophilic i.e. exhibit heat tolerance greater than
 CC normal proteins. In the case of kanamycin resistance, it allows
 CC recombinant microbes, e.g. *Thermophilus* species, to then be grown in the
 CC presence of the antibiotic at high temperatures. (updated on 17-OCT-2003
 CC to standardise OS field)

XX Sequence 1707 BP, 180 A, 581 C, 607 G, 339 T, 0 U, 0 Other;

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:	Mismatches:	Indels:	Gaps:
109.50	36.92%	1707	81	39	113	19
Best Local Similarity:	24.92%					
Query Match:	7.49%					

US-09-087-136-1 (1-275) x AAV32780 (1-1707)

```

Qy      2 SerGluLeuAapPro-----LeuAlaGluPhe-----Leu 11
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1137 GCGAAGAGAGACCGGCTGAGCTCCCTCCCTCCAGACACACACACCTCCGCCACCTC 1078

Qy      12 LeuProGluAapGlyAapArgAsnAlaArgGlnAsnAapProLeu-----26
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1077 CTCCCGAGGCGGCGC-----CGGAGAGAGACCGGCTTCCCTCCACCTCGGG 1030

Qy      27 -----1LeSerGlyGlyProLeu-----32
      |||:|||||:|||||:|||||:|||||:|||||:
Db      1029 GCTTTCAGGTGACGTAAGCCATCGGGGGGGGGGAGAGAGGCGAAGCTCGAGG 970

Qy      33 -----ProLeuGluSerProSerArg--LysLeuThrSer 43
      |||:|||||:|||||:|||||:|||||:|||||:
Db      969 AGCCTCGGGGCTCTCTCCGAGAGGGGCTTAAGACCTCTCGAGGCTCGGGAGAGCTCC 910

Qy      44 LeuLeuSerTyraaProThrVal-----51
      |||:|||||:|||||:|||||:|||||:|||||:
Db      909 CCCTTGAGCCACCGGGCGACGCTCTCCCTCGGGCTCTCTCGGGCGGCGAGGCGCAGG 850

Qy      52 ProGluSerProAapMetLysPheAlaArgLysArgLeuGlyAsnLeuLeuThrThrIle 71
      |||:|||||:|||||:|||||:|||||:|||||:
Db      849 GAGAAAGGCGCGAGAGAGAGGCGAGGAGGCGACCTTAACCTCTCTC-----799

Qy      72 LysHisIlePro-----SerGluIleIleGlyValLeuProGluAapTyr 86
      |||:|||||:|||||:|||||:|||||:|||||:
Db      798 CAAAGGCGACCCCTCCCGGAGAGGGGCGAGGCGCTCTCGGAGTACTG-----751

Qy      87 ThrArgAlaAapGluGluProGlyArgGlnGlyArgProProGlyArgProArgLysMet 106
      |||:|||||:|||||:|||||:|||||:|||||:
Db      750 GGGAACTCCAAAGCGAGAGAGGCGCGAGAGGCGACCGCGAAGCGCGCTCGCCCGCT 691

Qy      107 ProArgHisGluSerSerThrSerLeuMetGluSerProArgLysThrMetThrArgAap 126
      |||:|||||:|||||:|||||:|||||:|||||:
Db      690 CCGGTCCACAGTGGCGCCCGACAGCC--CACCAAAACCGCGCGGCTCGCGCGCGCG 634

Qy      127 SerLysIleMetPheGluLeuArgGly-----135
      |||:|||||:|||||:|||||:|||||:|||||:
Db      633 GATGCGCTTGGGAGCGCTCTCCCGGTGTACTACACGCTCTCCACCGCGAAGACGCTCG 574

Qy      136 -----LysProPheGluMetIleAlaGlyArgPheGluGluTyrSer 150
      |||:|||||:|||||:|||||:|||||:|||||:
Db      573 CCCGGGTTCACGCGTGGAGCGCTTGGGAGTGCATCTCTCCCTT--CGTCTCTCTCG 517

Qy      151 LeuGlyArgAlaTyrValIleGlyHisMetAsnAsnGluTyrGluProIleLysAlaGln 170
      |||:|||||:|||||:|||||:|||||:|||||:
Db      516 GCGAAGAGGCGCGCGGCGCGAGGCGCTTGGCGACACGAGAGAAAG-----472

Qy      171 ArgThrAapTyrAlaProAsnLeuAlaValAapTyrIleAlaCysAaGluIleHisArg 190
      |||:|||||:|||||:|||||:|||||:|||||:
Db      471 -----GACCAAGCGCGCGGATCGCGGCGCAT-----GACGAAGTCCAC--433

Qy      191 MetProArgProAapLysSerIleProGluLeuProIleValProSerArgIleAapIle 210

```

```

Db      432 CTCTCTCAAAAGGCTTCCAAAGGCTCCCGACGCGCTCGGCGT-----AAA 380
      |||:|||||:|||||:|||||:|||||:|||||:
Qy      211 PheAspAlaThrValAapProArgTyrGluThrAspLeuLysAsnGluTyrIleArgHis 230
      |||:|||||:|||||:|||||:|||||:|||||:
Db      379 GGGGATCTCG-----AGAGAGGCGCGCGCACCTGCAAA-----348

Qy      231 TrpLysGlnValLysLysGlyTyrCysAlaHisGlnArgArgArgThrAlaProHisAla 250
      |||:|||||:|||||:|||||:|||||:|||||:
Db      347 -----AAGAGAGGAGAGTGCATCCCGAGCGCAGAGAGAGTGGCTCTTAGA 300

Qy      251 ArgSerIleAlaLeu 255
      |||:|||||:|||||:|||||:|||||:
Db      299 AGAGCGCGCGCTCTC 285

RESULT 3
ABA92233/C
ID ABA92233 standard; DNA, 1707 BP.
XX
AC ABA92233;
XX
XX 29-AUG-2003 (revised)
DT 06-JUN-2002 (first entry)
DT
XX
DE Thermus flavus pyra gene.
XX
KM Pyra gene; dehydrogenase; enzyme; selectable marker; thermostable enzyme;
KM thermogenetics; gene; ds.
XX
OS Thermus thermophilus.
XX
FH
FT Key location/Qualifiers
FT CDS 262..810
FT /*tag= a
FT /product= "Dehydrogenase"
XX
PN US6350591-B1.
XX
PD 26-FEB-2002.
XX
XX 16-FEB-1999; 99US-00250585.
XX
PF 24-JUN-1994; 94US-00265522.
PR 30-JUN-1995; 95US-00496932.
PR 18-AUG-1997; 97US-00912794.
XX
XX (THER-) THERMOGEN INC.
XX
PI Weber JM, Demitjian DC, Casadaban MJ, Vonstein V, Pagratias NC;
XX WPI; 2002-279436/32.
XX
DR P-PSDB; AAM51136.
XX
XX Recombinant DNA comprising a DNA fragment isolated from Thermus flavus
PT and an exogenous DNA sequence that inserts into a Thermus chromosome
PT useful as a selectable marker for thermogenic selection.
XX
XX Example 3; Col 21-24; 23pp; English.
XX
XX The present sequence is that of the pyra gene of Thermus flavus strain
XX AT82. The gene encodes a thermostable dehydrogenase (see AAM51136). The
XX pyre gene was cloned from genomic DNA following construction of a gene
XX library and screening of recombinant cosmids for their ability to
XX transform a low-reversion Ura- strain of T. flavus to prototrophy. The
XX invention provides a new gene transfer system for extreme thermophiles of
XX the genus Thermus. A method of inserting an exogenous DNA sequence into a
XX targeted chromosomal region of a thermophilic microorganism involves:
XX constructing a plasmid vector which contains a targeting DNA sequence
XX that corresponds to the targeted chromosomal DNA; inserting an exogenous
XX DNA into the targeting DNA sequence so that it interbreeds with the normal
XX expression of the targeting sequence; transforming a host cell with the
XX vector; selecting stable transformants by screening at temperatures above
XX 55 degrees C; and confirming gene replacements by Southern blotting. The

```

CC pyA gene can be used as a selectable marker for specific chromosomal
CC integration in *Thermus*. The method can be used in thermogenetic processes
CC to generate thermostable enzymes and proteins for industrial processes.
CC (Updated on 29-AUG-2003 to standardise OS field)

Sequence 1707 BP; 180 A; 581 C; 607 G; 339 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2.85	Length:	1707
Score:	109.50	Matches:	81
Percent Similarity:	36.92%	Conservative:	39
Best Local Similarity:	24.92%	Mismatches:	94
Query Match:	7.49%	Indels:	113
DB:	6	Gaps:	19

US-09-087-136-1 (1-275) x ABA92233 (1-1707)

QY	2	SerGluIleAspPro-----LeuAlaGluPhe-----	Leu 11
Db	1137	GCGAAGTAGACCCCGCTGGCTCCCGCCCTTCACGACCAACCACTCCCACTC	1078
QY	12	LeuProGluAspGlyAspArgAsnAlaArgIlnAsnAspProLeu-----	26
Db	1077	CTCCCGGAGGGCGGC-----CGGGAAGAGACCCGGTTCCCTCCACTCGGG	1030
QY	27	-----IleSerGlyGlyProLeu-----	32
Db	1029	GCTTCCAGGTGACGGTAGCCATCCGGGGGGCGGGGGAAGGGCGAACGCTCGAG	970
QY	33	-----ProLeuGluSerProSerArg---LysIleuThrSer	43
Db	969	AGCCTCGGGGCTCTCTCCGAGGGGCGCTTAAGAACCTCTCGAGGCTCGGGGAAGACTCC	910
QY	44	LeuIleuSerTyzAspProThrVal-----	51
Db	909	CCCTGAGCCACCGGGCGACGAGTCTCTCCGCGCTTCCTCGGGCGCGAGGCGCAG	850
QY	52	ProGluSerProAspMetLysPheAlaArgLysArgLeuGlyAsnLeuLeuThrThrIle	71
Db	849	GAGAAAGGCGCGAGAAGAGGGCGAGCGAGAGCGACCCCTAAACTCTCTC-----	799
QY	72	LysIleHisAspPro-----SerGluIleIleGlyValLeuProGluAspTyr	86
Db	798	CAAGGGCACCCCTCCCGGCGAGGGGGCAGCGCTCTCGGAGATACG-----	751
QY	87	ThrArgAlaAspGluGluProGlyArgGlnIlyArgProProGlyArgProArgLysMet	106
Db	750	GCGAAACCTCCAGGCGAGGAGGGGGCGGAGAGGCAAGCCAGAGGCCCGCTCGCCCCGCT	691
QY	107	ProArgHisGluSerSerThrSerLeuMetGluSerProArgLysThrMetThrArgAsp	126
Db	690	CCGGTCCACATGGAGCCCCACAGCC---CACCAAAACCCCGCCGGGCTCCGCGCGCG	634
QY	127	SerIlysiIleMetPheGluLeuArgGly-----	135
Db	633	GATGGCTTGGGAGCGCTCTCCCGGCTGTTACACAGCTCTCCACGCGCAAGAGCTCG	574
QY	136	-----LysProPheGluMetIleAlaGlyArgPheGluGluGlyTyrSer	150
Db	573	CCCGGATTCAAGGTAGACCCCTTCGAGATAGCATCCCTCCCTT---CCGTCCTTCG	517
QY	151	LeuGlyArgAlaIleValIleGlyLysIleMetAsnAsnGluTyrGluProIleLysValAlaGln	170
Db	516	GCGAAGAGGCGCGCGCGCCAGGCTTCGCGCACCAAGAGAAAG-----	472
QY	171	ArgThrAspTyzAlaProAsnLeuAlaValAspTyrLeuAlaCysArgGluIleHisArg	190
Db	471	-----GACACCGCCCCCGATGGCGGGGCGAT-----GACGAAAGTCCAC---	433
QY	191	MetProArgProAspLysSerIleProGluLeuProIleValProSerArgGluIleAspGln	210
Db	432	CTTCT---CGTCTCAAAAGCTTTTCCAAAGCTCTCCACAGGCTCGGCT-----AAA	380

Qy	211	PheAspAlaThrValAspProAlaGlyTrgAutThrAspLeuLysAsnGluTrpIleArgHis	230		
Db	379	GGGAGTCTGG-----AGGAGGCCGCCGACTGCAGAA-----	348		
Qy	231	TrpLysGlnValLysLysGlyTrpCysAlaHisGlnArgArgArgThrAlaProHisAla	250		
Db	347	-----AAGAAGGGGGAGTGTCAATCCCGACGCGAGAGGAGAAAGTGGCCCTTAAG	300		
Qy	251	ArgSerIleAlaLeu	255		
Db	299	AGAGCCCCGCTCTC	285		
RESULT 4					
AAC6404					
ID	AAC6404	standard; DNA; 1315 BP.			
XX					
AC	AAC6404;				
XX					
DT	18-OCT-2000	(first entry)			
XX					
DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 50014.				
XX					
KW	Hybridization assay; genetic mapping; gene expression control;				
KW	protein identification; signal transduction pathway; metabolic pathway;				
XX	promoter; termination sequence; ss.				
XX					
OS	Arabidopsis thaliana.				
PN	EP1033405-A2.				
PD					
XX					
PF	06-SEP-2000.				
XX					
XX	25-FEB-2000;	200BP-00301439.			
PR	25-FEB-1999;	99US-0121825P.			
PR	05-MAR-1999;	99US-0123180P.			
PR	09-MAR-1999;	99US-0123548P.			
PR	23-MAR-1999;	99US-0125788P.			
PR	25-MAR-1999;	99US-0126264P.			
PR	29-MAR-1999;	99US-0126785P.			
PR	01-APR-1999;	99US-0127462P.			
PR	06-APR-1999;	99US-0128234P.			
PR	08-APR-1999;	99US-0128714P.			
PR	16-APR-1999;	99US-0129845P.			
PR	19-APR-1999;	99US-0130077P.			
PR	21-APR-1999;	99US-0130449P.			
PR	23-APR-1999;	99US-0130510P.			
PR	28-APR-1999;	99US-0130891P.			
PR	30-APR-1999;	99US-0131448P.			
PR	30-APR-1999;	99US-0132048P.			
PR	30-APR-1999;	99US-0132407P.			
PR	04-MAY-1999;	99US-0132484P.			
PR	05-MAY-1999;	99US-0132485P.			
PR	06-MAY-1999;	99US-0132486P.			
PR	06-MAY-1999;	99US-0132487P.			
PR	07-MAY-1999;	99US-0132863P.			
PR	11-MAY-1999;	99US-0134256P.			
PR	14-MAY-1999;	99US-0134218P.			
PR	14-MAY-1999;	99US-0134219P.			
PR	14-MAY-1999;	99US-0134370P.			
PR	18-MAY-1999;	99US-0134376P.			
PR	19-MAY-1999;	99US-0134941P.			
PR	20-MAY-1999;	99US-0135124P.			
PR	21-MAY-1999;	99US-0135353P.			
PR	24-MAY-1999;	99US-0135629P.			
PR	25-MAY-1999;	99US-0136021P.			
PR	27-MAY-1999;	99US-0136392P.			
PR	28-MAY-1999;	99US-0136782P.			
PR	01-JUN-1999;	99US-0137222P.			
PR	03-JUN-1999;	99US-0137528P.			
PR	04-JUN-1999;	99US-0137502P.			
PR	07-JUN-1999;	99US-0137724P.			

PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 21-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145086P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145216P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 03-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.

PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160814P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160982P.
PR 22-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:
Pred. No.:
Score:
Percent Similarity:

2.29
109.00
36.93%

Length:
Matches:
Conservative:

1315
69
37

Best Local Similarity: 24.04%
 Query Match: 7.46%
 DB: 3
 Indels: 82
 Gaps: 16
 US-09-087-136-1 (1-275) x AAC46404 (1-1315)

QY 5 AsppPro-----LeuAlaGluPheLeuLeuProGluAspGlyAspArgAsnAlaArg 21
 DB 321 GATCCGTCAGCTGTTCTTTACCCGTTCTCTACCGGGTCAGGG---TTTTCAGCCGA 377
 QY 22 Gln-----AsnAspProLeuLeuSerGlyGlyProLeuProLeuGluSer 36
 DB 378 CCTGTCGAGGATTCGTCGAGATCTTCCTCCGTCAGCGCTGATCTGAGGGGTACCT 437
 QY 37 ProSerArgGlySerLeuThrSerLeuLeuSerGlyThrAspProThrValProGluSerProAsp 56
 DB 438 CCTGGCCCGAGTTTACT-----TACGATCTCGACCACTATGACGCGTCA 485
 QY 57 Met-----LysPheAlaArgLysArgLeuGlyAsnLeuLeuThrThrIle 71
 DB 486 ATGGAATCTTGTTCAGCAGATTATATAGGAGAGAAATCTCAGATT----- 533
 QY 72 LysHisHisProSerGluIleIleGlyValLeuProGluAspGlyThrArgAlaAspGlu 91
 DB 534 -----CGTCAATTACCAAGCCTTGATG----- 557
 QY 92 GluProGluArgGlnGlyArgProProGly----- 101
 DB 558 -----GGTCTCTCTGTTGGTGGTCCCTTAGACCAAGTCTCAATT 602
 QY 102 ---ArgProArgGlyMetProArgHisGluSerSerThrSerLeuMetGluSerPro--- 119
 DB 603 CTGCAACCGCGGGTGGCCCG-----CCCCCAACTCAATCTAGACACTGACAG 653
 QY 120 ---ArgLysThrMetThrArgAspSerLysIleMetPheGluLeuArgLysProPhe 138
 DB 654 AATAGAAAGCCAGAGCAAGATGGCGCTCTTGGCGTT---GTTAGAGGGAGAAAGTTC 710
 QY 139 GluMetIleAlaGlyArgPheGluGluGluGlySerLeuGlyArgAlaThrValIleGly 158
 DB 711 CGATACCTAGAGGGGTCT---TCTTCTTTATTATTCCTTGTCGATCGTGGTGA--- 764
 QY 159 HisMetAsnAsnGluIuArgLupProIleLysAlaGlnArgThrAspThrAlaProAsnLeu 178
 DB 765 ---AATGTCGCTCAGCTTGAATTCAGCCGCAAGAGAGTGTATA----- 806
 QY 179 AlaValAspTyrLeuAlaCysArgGluIleHisArgMetProArgPro----- 194
 DB 807 -----ATGAAACCTTTGCCAAACCGTTACTCTGTGAT 839
 QY 195 -----AspLysSerIleProGluLeuProIleValProSerArgIleAsp----- 209
 DB 840 TTAAACAACGAGACAGATGTACCATGATCATGATGAAGATCAGCGATGAGACAA 899
 QY 210 GluPheAspAlaThrValAspProArgTyrGluThrAspLeuLysAsnGluTyrIleArg 229
 DB 900 GAGGACGAGGAAGCTGTAAAGCAATTATCTGAAAGATCTTTGAAAAGACACATGAG 959
 QY 230 HisTPrpGlnValIleLysGlyGlyTyrCysAlaHisGlnArgArgTThrAlaProHis 249
 DB 960 CGAGCTAAGAAAGTCGTCGCAATTTACGAGAAAGACGTCGAGAGATCAGAGGTAC 1019
 QY 250 AlaArgSerIleAlaLeuIle 256
 DB 1020 AAAGAAAGATTAACCTTATT 1040

RESULT 5
 AAC39522
 ID AAC39522 standard; DNA; 1318 BP.
 AC AAC39522;
 XX
 XX 17-OCT-2000 (first entry)

XX XX Arabidopsis thaliana DNA fragment SEQ ID NO: 24943.
 DE Hybridisation assay; genetic mapping; gene expression control;
 XX protein identification; signal transduction pathway; metabolic pathway;
 KW promoter; termination sequence; ss.
 OS Arabidopsis thaliana.
 PN EP1033405-A2.
 XX PD 06-SEP-2000.
 XX PF 25-FEB-2000; 2000EP-00301439.
 XX PR 25-FEB-1999; 99US-0121825P.
 PR 05-MAR-1999; 99US-0123180P.
 PR 09-MAR-1999; 99US-0123548P.
 PR 23-MAR-1999; 99US-0125788P.
 PR 25-MAR-1999; 99US-0126264P.
 PR 29-MAR-1999; 99US-0126785P.
 PR 01-APR-1999; 99US-0127462P.
 PR 06-APR-1999; 99US-0128234P.
 PR 08-APR-1999; 99US-0128714P.
 PR 16-APR-1999; 99US-0129845P.
 PR 19-APR-1999; 99US-0130077P.
 PR 21-APR-1999; 99US-0130449P.
 PR 23-APR-1999; 99US-0130510P.
 PR 28-APR-1999; 99US-0130891P.
 PR 30-APR-1999; 99US-0131449P.
 PR 30-APR-1999; 99US-0132048P.
 PR 04-MAY-1999; 99US-0132407P.
 PR 05-MAY-1999; 99US-0132484P.
 PR 06-MAY-1999; 99US-0132485P.
 PR 07-MAY-1999; 99US-0132486P.
 PR 07-MAY-1999; 99US-0132487P.
 PR 11-MAY-1999; 99US-0132633P.
 PR 14-MAY-1999; 99US-0134256P.
 PR 14-MAY-1999; 99US-0134218P.
 PR 14-MAY-1999; 99US-0134219P.
 PR 14-MAY-1999; 99US-0134221P.
 PR 14-MAY-1999; 99US-0134370P.
 PR 18-MAY-1999; 99US-0134768P.
 PR 19-MAY-1999; 99US-0134941P.
 PR 20-MAY-1999; 99US-0135124P.
 PR 21-MAY-1999; 99US-0135124P.
 PR 24-MAY-1999; 99US-0135353P.
 PR 25-MAY-1999; 99US-0135629P.
 PR 27-MAY-1999; 99US-0136021P.
 PR 28-MAY-1999; 99US-0136392P.
 PR 01-JUN-1999; 99US-0136782P.
 PR 03-JUN-1999; 99US-0137222P.
 PR 04-JUN-1999; 99US-0137528P.
 PR 07-JUN-1999; 99US-0137502P.
 PR 08-JUN-1999; 99US-0137724P.
 PR 10-JUN-1999; 99US-0138094P.
 PR 10-JUN-1999; 99US-0138540P.
 PR 14-JUN-1999; 99US-0138847P.
 PR 16-JUN-1999; 99US-0139119P.
 PR 17-JUN-1999; 99US-0139452P.
 PR 18-JUN-1999; 99US-0139453P.
 PR 18-JUN-1999; 99US-0139454P.
 PR 18-JUN-1999; 99US-0139455P.
 PR 18-JUN-1999; 99US-0139456P.
 PR 18-JUN-1999; 99US-0139457P.
 PR 18-JUN-1999; 99US-0139458P.
 PR 18-JUN-1999; 99US-0139459P.
 PR 18-JUN-1999; 99US-0139460P.
 PR 18-JUN-1999; 99US-0139461P.
 PR 18-JUN-1999; 99US-0139462P.
 PR 18-JUN-1999; 99US-0139463P.
 PR 18-JUN-1999; 99US-0139750P.
 PR 18-JUN-1999; 99US-0139763P.

PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144086P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 21-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145086P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.

PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-015753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161929P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:
Pred. No.: 2.3 Length: 1318
Score: 109.00 Matches: 69
Percent Similarity: 36.93% Conservative: 37
Best Local Similarity: 24.04% Mismatches: 99
Query Match: 7.46% Indels: 82
DB: 3 Gaps: 16

US-09-087-136-1 (1-275) x AAC39522 (1-1318)

QY 5 AspPro-----LeuAlaGluPheLeuLeuProGluAAspArgAspAlaArg 21
DB 321 GATCGTACGCTGTTCTTATACCCGTTGCTTACCGGCTGAGGG---TTTTCACCCCGA 377
QY 22 Gln-----AsnAspProLeuIleSerGlyGlyProLeuProLeuGluSer 36
DB 378 CCGTGTGAGGATTGCTGCGACGATCCTTCGCTGACGGCTGGTAATCTGATGGGTAACCT 437
QY 37 ProSerArgIleuThrSerLeuLeuSerTyrAspProThrValProGluSerProasp 56
DB 438 CCTGCGCCGAGTTTACT-----TACATCTTGACCATATGACGACGCTCA 485
QY 57 Met-----LysPheAlaArgIleValArgLeuGlyAsnLeuLeuThrThrIle 71

Db	486	ATGGAATCCTTGTTCGACGAGTTTAAAGGGAGAGAAATCTCAAGATT-----	533
Qy	72	Lysh1sh1sProSeRglu1le1leGlyValleuProGluAspTyrThrArg1a1aSpGlu	91
Db	534	-----CGTCCATTACCAAGCCCTTGATTG-----	557
Qy	92	GluProG1yArgG1uG1yArgProG1y-----	101
Db	558	-----GGTTCTCTCTGTGGGTGGGTGCCATTAGACCAAGTCTCAATT	602
Qy	102	---ArgProArgLyMetProArgHisGluSerSerThrSerLeuMetGluSerPro---	119
Db	603	CTGCACCGCGGGGTGGCCCG-----CCCCCAACTTCATTTCTAGACCTAGCAGG	653
Qy	120	---ArgLyThMetThrArgAspSerLyS1leMetPheGluLeuArgG1yLysProPhe	138
Db	654	AATAGAAAGCCAGAAAGCAGATGGGCTCTGGCCGTT---GTTAGAGGAGAAAGGTC	710
Qy	139	GluMet1leAlaG1yArgPheGluGluGluTyrSer1euG1yArgAlaTTrpValLySg1y	158
Db	711	CGAATCAGTGAAGGGGTCT---TCTTCTCTTATTCATTGTCATCGTGTTGAAA---	764
Qy	159	HisMetAenAsnGluTyrGluPro1leLySAlaG1uArgThraAspTyrAlaProAsnLeu	178
Db	765	-----AATGGTGTCTACGTTGGAATTCAGCCGCAAAAGAGGTGATA-----	806
Qy	179	AlaValAspTyrLeuAlaCysArgGlu1leHisArgMetProArgPro-----	194
Db	807	-----ATGAACCTTTGGCCAAACCGTTACTGTGAT	839
Qy	195	-----AspLySer1leProGluLeuPro1leValProSerArg1leAsp-----	209
Db	840	TTAACACAGACAGACAAAGTACAGATGATCCAGATGACATGAAGATCAGCGGATGAAGACAA	899
Qy	210	GluPheAspAlaThValAspProArgTyrGluThraPheLeuLysAsnGluTyr1leArg	229
Db	900	GAGGACGAGGAAGCTGTAAAGCAATATCTGAGGAAGATCTTTGMAAAGACACATCGAG	959
Qy	230	HisTrpLysG1uValLyLysG1yTrpCysAlaHisG1uArgArgArgThraAlaProHis	249
Db	960	CGAGCTTAAGAAAGTCCGTGCACAAATTACGAGAAAGACGTCACAGAGATCAGAGGTAC	1019
Qy	250	AlaArgSer1leAlaLeu1le	256
Db	1020	AAAGAAAGATTAACCTTATT	1040
RESULT 6			
ABQ15200/c			
Db	ID	ABQ15200 standard; DNA; 1428 BP.	
XX	AC		
XX	ABQ15200;		
XX	12-JUL-2002 (first entry)		
DE	Oligonucleotide for detecting cytosine methylation SEQ ID NO 1791.		
KM	Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;		
KM	drug; side effect; cancer; central nervous system; cardiovascular;		
KM	gastrointestinal; respiratory system; single nucleotide polymorphism;		
XX	SNP; cell differentiation; ds.		
OS	Homo sapiens.		
XX	WO200218632-A2.		
XX	07-MAR-2002.		
XX	01-SEP-2001; 2001WO-EP010074.		
XX	01-SEP-2000; 2000DE-01043826.		
XX	05-SEP-2000; 2000DE-01044543.		

PA (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Guetig D;

AA WPI; 2002-371829/40.
DR

PT Determining the degree of cytosine methylation in genomic DNA, useful for diagnosis and prognosis, comprises selective hybridization of amplicons from chemically treated DNA.

AA Claim 12; 56pp + Sequence Listing; 56pp; German.
PS

CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g., cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABO1310-
CC ABO54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention

Sequence 1428 BP; 194 A; 184 C; 630 G; 420 T; 0 U; 0 Other;

Alignment Scores:

Pred. NO.:	3.49	Length:	1428
Score:	107.50	Matches:	74

Percent Similarity:	35.33%	Conservative:	38
Best Local Similarity:	23.34%	Mismatches:	113

Query Match:	7.35%	Indels:	93
DB:	6	Gaps:	15

ITS-09-087-136-1 (1-275) x AB015200 (1-1428)

On 12 Feb 2006, the following data were collected:

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841

1100

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	52
--	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

[illegible][illegible]

27

-----GCGCGCCLCGACGACGAAACCLCGCAACCG
10/8

5 oserginelliegylvalleuprocinaspipyrinargalala
 ...
 ...

D5 1029 GACC-----CCGACCCCGACTCGACTCGCG

```

95 ggInglYargProProGlyArgProArgLysMetProArgHisG
QY ||||| ... |||

```

Db : 992 -----GAACCCCCCGAACCGACCGCGACGTCGAAAAAACATA

```

115 wnetGluserProarg-----lysthmett

```

Db 939 AAAAAAAAAATACGAAAAACAACATCCCGAATTCAAAACGAAAA


```
QY 161 naangluTyrgluProIlelysaIaGlnArgThraSPYrAlaProAsnLeuAlaValAs 181
CC |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
CC CGAAGCGCAGATCTTAAAAAATACGCGCCGACCGCCGACCGCCGCGCCGCGC-- 787
Db 730 CGAAGCGCAGATCTTAAAAAATACGCGCCGACCGCCGACCGCCGCGCGCCGCGC-- 787
QY 181 pTyrlauAlaCyArggluIleHisArgMetProArgProAspLySerIlePro----- 199
CC |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
CC 788 -----CAACCCCGCACCGCTCCGCAAAACCGCCCGCGCGCGCCGCGCCCTCC 834
Db 200 -----GluLeuProIleValProSerArgIleAspGluPh 211
QY 835 AAACCGACCTCCGCTTAACGAAACCGCCGCGCGCGCGCGCGCGCGCGCGACT 891
Db 211 eaSPaIaThraValaSPProArgTyrgluThraSPleuLyAsnGluTyrlleArgHisTr 231
CC |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
CC 892 CGACGGGACCT--TCGCCCCCTTATAAAAAAACTTTCTTTAAACAGATCCGCTAAAA 948
QY 231 pTyrgluIle-----VallylsyglTyTrpCysAlaHisGlnArgArgArgTh 246
CC |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
CC 949 CGGAACCTTCACCCCGACCTTTTACAAACAAACCTTCGACGAAACGCTCGACCGCGAAC 1008
QY 246 rAlaProHisAlaArgSerIleAlaLeuIleAsnlySileTyrgluPro 262
Db 1009 GACCCGCG-----AAATCATGCTCTTAAAAAATATTATTATTTCCCG 1051
RESULT 8
ABN80010/c
ID ABN80010 standard; DNA; 6906 BP.
AC ABN80010;
DT 15-JUL-2002 (first entry)
XX
DE Human chemically modified disease associated gene SEQ ID NO 27.
XX
KW Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;
KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;
KW antidiabetic; cyostatic; anticonvulsant; de.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200200927-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007536.
XX
PR 30-JUN-2000; 2000DE-01032529.
XX
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130908/17.
XX
PT Novel nucleic acid useful for diagnosis and therapy of diseases
PT associated with development genes such as diabetes, comprises a sequence
PT of a segment of chemically pretreated DNA of genes associated with
PT development.
XX
PS Claim 1; SEQ ID NO 27; 27bp; English.
XX
CC The invention relates to a nucleic acid (I) comprising a sequence at
CC least 18 bases in length of a segment of chemically pretreated DNA (II)
CC of genes associated with development selected from 87 genes listed in the
CC specification such as ACCPN, ADPN, or APDI and comprising one of 350
CC sequences (ABN79984-ABN80333) or their complements. The invention is
CC useful for the diagnosis or therapy of diseases associated with
CC development genes, in particular disease related to homeobox containing
CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes
```

```
CC associated with congenital heart disease, epilepsy, diseases related to
CC histone deacetylation, Currarino syndrome, diseases related with the
CC development of the brain and limb girdle muscular dystrophy and dwarfism.
CC Oligomers specific to each of the genes are useful for detecting the
CC methylation state of all CpG dinucleotides within the 350 sequences or
CC (II) and their complementary sequences, as primer oligonucleotides for
CC the amplification of the 350 sequences, (II) and/or their complements
CC as oligomer probes for detecting the cytosine methylation state and/or
CC single nucleotide polymorphisms (SNPs). Note: The sequence data for this
CC patent did not form part of the printed specification but is based on
CC sequence information supplied to Derwent by the European Patent Office
SQ Sequence 6906 BP; 1343 A; 390 C; 2193 G; 2980 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 24.8 Length: 6906
Score: 107.50 Matches: 74
Percent Similarity: 35.33% Conservative: 38
Best Local Similarity: 23.34% Mismatches: 113
Query Match: 7.35% Indels: 93
DB: Gaps: 15
US-09-087-136-1 (1-275) x ABN80010 (1-6906)
QY 12 LeuProGluAspGlyAspArgAsnAlaArgGlnAsnAspProLeuIleSerGlyPro 31
Db 3045 CTACCCGACAAACCCCGCAAAACGCTAAAGCGCG -AATCCGCGCGCTCCCTCCGCGCG 2987
QY 32 LeuPro-----LeuGluSerProSerArg----- 39
Db 2386 ACCCCTAACGCGCCCGCAAAATTAACGACTTACCAACTTCAAAAAACCGCCCGACATCA 2927
QY 40 -----LysLeuThrSerLeuLeuSer-TyrAspProThrValProGluSerPr 55
Db 2326 AAACGCGCATCCGCGCCGACCTCGACTCGACCTCGACCTCACTCACTTATCCCTCC----- 2875
QY 55 oAspMetLyPheAlaArgLysArgLeuGlyAsnLeuLeuThrThrIleLysHisSpr 75
Db 2874 -----CGGCGCGCTGACGACGAAACCTGCAAAAGCCCTTAACAAAAACC 2828
QY 75 oSerGluIleIleGlyValLeuProGluAspTyrrThraArgAlaAspGluGluTyar 95
Db 2827 GACC-----CCGACCCCGCACTCGACTCGGACGACGACAAACCC----- 2791
QY 95 gGlnGlyArgProProGluTyArgProArgLysMetProArgHisGluSerThrSerIle 115
Db 2790 -----GAAACCCCGCAACGACGCGACGCGCAAAACATTAATAAAATCGACGCA 2738
QY 115 uMetGluSerProArg-----LysThrMetThrArgAsp----- 126
Db 2737 AAAAAAATACGAAAAAACAATCCCGAATTCAAAAAGAAAAACGACCGAAAAA 2678
QY 127 -----SerLySileMetPheGluLeuArgGlyLy 136
Db 2677 AAACGACGCGCGCGCGCGCTTTTATACAAAAATCGAAGAACTCCGCGCGCG 2618
QY 136 sPro-----PheGluMetIleAlaGlyArgPheGlu----- 146
Db 2617 CCGACCGCATTTTCAACGCAATCTTATTCACAAATCTCGACGACCAAAATATCC 2558
QY 147 -----GluGluTySerLeuGlyArgAlaTrpValIleGlyHisMetAs 161
Db 2557 TCGAAGCGTTCGACAAACAACTACGCTAATTAATTAACATACCTTCAAAACCGGA 2498
QY 161 naangluTyrgluProIlelysaIaGlnArgThraSPYrAlaProAsnLeuAlaValAs 181
Db 2497 CGAAGCGCAGATCTTAAAAAATACGCGCCGACCGCCGACCGCCGCGCGCGC-- 2440
QY 181 pTyrlauAlaCyArggluIleHisArgMetProArgProAspLySerIlePro----- 199
Db 2439 -----CAACCCCGCACCGCTCCGCAAAACCGCCCGCGCGCGCCCTCC 2393
QY 200 -----GluLeuProIleValProSerArgIleAspGluPh 211
```

```
Db 2392 AACCCGACCTCCGCTAACCGAACCCCTCCCG---CCCCCGGCGCCACGCGCGACT 2336
Qy 211 eAspAlaThyValAspProAlaGlyTrpGluThraSpleuYsaAngluYrIleAghHisTr 231
Db 2335 CGAGACTTCAC---TCGCCCCCTATMAAAMAACTTTCTTTAAACGACTCCGCTAATA 2279
Qy 231 pLyseGln-----ValLyseGlyTrpCyAlaHisGlnAraGArGArGTh 246
Db 2278 CCGAACTTCACCCCGACCTTTTACACAAACCTTCAGACGAGCGCTCGACCGCGAAC 2219
Qy 246 rAlaProHisAlaAraSerIleAlaLeuIleAsnLyseIleYrGlnPro 262
Db 2218 GACCCCG-----AAATCCATCGCTCTAATAATATTTATTTTCCCG 2176

RESULT 9
ADG75175
ID ADG75175 standard; DNA; 9369 BP.
AC ADG75175;
XX
XX 11-MAR-2004 (first entry)
XX
XX Human herpesvirus 2 strain HG52 UL36 DNA - SEQ ID 247.
XX
XX immunogenic; herpes simplex virus type 2; HSV; virucide; HSV2 infection;
XX vaccine; ds; UL36; gene.
XX
XX Human herpesvirus 2 strain HG52.
XX
XX MO2003086308-A2.
XX
XX 23-OCT-2003.
XX
XX 09-APR-2003; 2003WO-US011231.
XX
XX 11-APR-2002; 2002US-00121988.
XX 19-JUL-2002; 2002US-00200562.
XX 06-SEP-2002; 2002US-00237551.
XX
XX (CORI-) CORIXA CORP.
XX
XX PA Day CH, Hosken NA, Parsons JM;
XX
XX PI
XX
XX DR WPI; 2003-845255/78.
XX
XX DR P-PSDB; ADG75178.
XX
XX PT New isolated polypeptide comprising at least an immunogenic portion of a
XX PT herpes simplex virus (HSV) antigen, useful for diagnosing, treating or
XX PT preventing HSV infection.
XX
XX PS Claim 3; SEQ ID NO 247; 383bp; English.
XX
XX
XX The invention relates to a novel isolated polypeptide comprising at least
XX CC an immunogenic portion of a herpes simplex virus (HSV) antigen. The
XX CC polypeptide of the invention demonstrates virucide activity and may be
XX CC useful for diagnosing, treating or preventing Human herpesvirus 2 (herpes
XX CC simplex virus type 2, HSV2) infection via the production of a vaccine.
XX CC The current sequence is that of an HSV2 DNA which was used in the
XX CC exemplification of the invention.
XX
XX SQ Sequence 9369 BP; 1179 A; 3772 C; 3163 G; 1255 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 40.3 Length: 9369
Score: 107.00 Matches: 62
Percent Similarity: 32.00% Conservative: 18
Best Local Similarity: 24.80% Mismatches: 93
Query Match: 7.32% Indels: 77
DB: 10 Gaps: 10

US-09-087-136-1 (1-275) x ADG75175 (1-9369)
```

```
Qy 13 ProGluAspGlyAraProAsnAlaArgGlnAsnAspProLeuIleSerGlyProLeu 32
Db 8063 CCGAGGGGTGGCGTCCCG-----CGGACATCCCCCGCGGTGGCCCGGCTTCCGCTT 8116
Qy 33 ProLeuGluSerProSerArgLyLeuThraSerLeuLeuSerTyraAspProThraValPro 52
Db 8117 CCGCGCTCTCTCCGACACAGCCG-----CCACCC 8146
Qy 53 GluSerProAspMetLysPheAlaArgLyAraGluLeuThraValLeuThraIleLys 72
Db 8147 CGAGATCCG-----CGCCCGCGGCTTGGTTTCGCTTTCGCTTCCGCTTCCG 8191
Qy 73 HisHisProSerGluIleIleGlyValLeuProGluAspTyraThraAlaAspGlu--- 91
Db 8192 CCGGCCCCGCTCGGCTCGCGGGGCTTTCGCGGCCCGACGCTCGCCCATCCCGCTC 8251
Qy 92 GluPro-GlyArgGlnGlyArgProProGlyArgProArgLyMetProArgHisGluSe 111
Db 8252 CGCCCGCGGAAGAGACGGTGTGCTCCCGGGGGGAGACCGCGCGCGCGCGACAA 8311
Qy 111 rSerThraSerLeuMetGluSerProArgLyThraMetThraArgAspSerLyIleMetPh 131
Db 8312 CAGCGGGTCCCTCGCGACACCCCGCGGGCGCGCGCGCGCGCGCGCGCTCA 8371
Qy 131 eGluLeuArgGlyLysProPheGluMetIleAlaGlyArgPheGluGluGlyTrsSerle 151
Db 8372 CGCGCCCGCGCGTCC-----CTCGCTCT 8395
Qy 151 uGlyArgAlaTrpValLyseGlyHisMetAsnAsnGluTyrgluProIleLys-AlaGlnA 171
Db 8396 CGGCTCGTT-----GAACTCCCTCCCTCCCGCGGAGC 8431
Qy 171 rGThraAspTyraAla-ProAsnLeuAlaValAspTyraLeuAlaCyAraGluIleHisArg 190
Db 8432 CAGCCGACACACCGCGCGGTTCGCGCGGAGCGGCGGCGCGCTTCGCGGCGC 8491
Qy 191 Met---ProArgProAspLySerIleProGluLeuProIleValProSerArgIleAsp 209
Db 8492 TTGCCCCCGACCTCGCGCGTCCAGACAGCC-----8524
Qy 210 GluPheAspAlaThraValAspProArgTyrgluThraSpleuYsaAngluYrIleAraG 229
Db 8525 -----CCCCGCTT-----8533
Qy 230 HisTrpLyseGlnValys-----LysGlyTrpCyAlaHis 241
Db 8534 ---TGCCCCGAGCGCGGTGCCCCGAGCGAGCCTTGTGCGTGTGGTCTGCGCGGAG 8590
Qy 242 GlnAraGArgThraAlaProHis 249
Db 8591 GCCCGTTGCCCGCGCCCGCCAC 8614

RESULT 10
ADG75118
ID ADG75118 standard; DNA; 9369 BP.
XX
XX AC ADG75118;
XX
XX 11-MAR-2004 (first entry)
XX
XX DE Human herpesvirus 2 UL36 ORF DNA - SEQ ID 190.
XX
XX immunogenic; herpes simplex virus type 2; HSV; virucide; HSV2 infection;
XX vaccine; ds; UL36; gene.
XX
XX Human herpesvirus 2.
XX
XX MO2003086308-A2.
XX
XX 23-OCT-2003.
XX
XX 09-APR-2003; 2003WO-US011231.
XX
```


FT /tag= k
FT /product= "ORF#6b protein"
FT /note= "encoded protein shown in AAW72245"
FT 6446. .8482
FT /tag= l
FT /product= "ORF#6a protein"
FT /transl_except= (pos: 7400. .7402, aa: Ala-Ala)
FT /transl_except= (pos: 7481. .7486, aa: Ile)
FT /note= "encoded protein shown in AAW72244"
FT 8457. .9347
FT /tag= m
FT /product= "ORF#7 protein"
FT /note= "encoded protein shown in AAW72175"
FT complement(9604. .11855)
FT /tag= n
FT /transl_except= (pos: 11635. .11636, aa: Ala)
FT /product= "ORF#8 protein"
FT /note= "encoded protein shown in AAW72176"
FT complement(11905. .14520)
FT /tag= p
FT /product= "ORF#9a protein"
FT /note= "encoded protein shown in AAW72223"
FT complement(11905. .14508)
FT /tag= o
FT /product= "ORF#9b protein"
FT /note= "encoded protein shown in AAW72222"
FT 14399. .15802
FT /tag= q
FT /product= "ORF#10 protein"
FT /note= "encoded protein shown in AAW72177"
FT complement(15996. .16286)
FT /tag= r
FT /product= "ORF#11 protein"
FT /note= "encoded protein shown in AAW72178"
FT complement(16202. .18064)
FT /tag= s
FT /product= "ORF#12 protein"
FT /note= "encoded protein shown in AAW72179"
FT complement(18105. .19661)
FT /tag= t
FT /product= "ORF#13 protein"
FT /note= "encoded protein shown in AAW72180"
FT complement(19415. .20074)
FT /tag= u
FT /product= "ORF#14 protein"
FT /note= "encoded protein shown in AAW72181"
FT 20155. .21453
FT /tag= v
FT /product= "ORF#15 protein"
FT /note= "encoded protein shown in AAW72182"
FT complement(21326. .22291)
FT /tag= w
FT /product= "ORF#16 protein"
FT /note= "encoded protein shown in AAW72183"
FT complement(22546. .24654)
FT /tag= x
FT /product= "ORF#17 protein"
FT /note= "encoded protein shown in AAW72184"
FT 24684. .25955
FT /tag= y
FT /product= "ORF#18 protein"
FT /note= "encoded protein shown in AAW72185"
FT complement(26295. .27251)
FT /tag= z
FT /product= "ORF#19 protein"
FT /note= "encoded protein shown in AAW72186"
FT complement(27630. .31784)
FT /tag= ab
FT /product= "ORF#20a protein"
FT /note= "encoded protein shown in AAW72224"
FT complement(27630. .31754)
FT /tag= aa
FT /product= "ORF#20b protein"

FT /note= "encoded protein shown in AAW72225"
FT complement(32067. .32735)
FT /tag= ac
FT /product= "ORF#21 protein"
FT /note= "encoded protein shown in AAW72187"
FT 33140. .34984
FT /tag= ad
FT /product= "ORF#22a protein"
FT /note= "encoded protein shown in AAW72226"
FT 33386. .34984
FT /tag= ae
FT /product= "ORF#22b protein"
FT /note= "encoded protein shown in AAW72227"
FT complement(35205. .37721)
FT /tag= af
FT /product= "ORF#23 protein"
FT /note= "encoded protein shown in AAW72188"
FT complement(38058. .39188)
FT /tag= ag
FT /product= "ORF#24 protein"
FT /note= "encoded protein shown in AAW72189"
FT 39090. .39935
FT /tag= ah
FT /product= "ORF#25 protein"
FT /note= "encoded protein shown in AAW72190"
FT 40216. .41973
FT /tag= ai
FT /product= "ORF#26 protein"
FT /note= "encoded protein shown in AAW72191"
FT 42206. .44178
FT /tag= aj
FT /transl_except= (pos: 44063. .44064, aa: Lys)
FT /product= "ORF#27 protein"
FT /note= "encoded protein shown in AAW72192"
FT complement(44853. .47297)
FT /tag= ak
FT /product= "ORF#28 protein"
FT /note= "encoded protein shown in AAW72193"
FT 47122. .47338
FT /tag= al
FT /product= "ORF#29 protein"
FT /note= "encoded protein shown in AAW72194"
FT complement(47305. .49662)
FT /tag= am
FT /product= "ORF#30 protein"
FT /note= "encoded protein shown in AAW72195"
FT complement(50035. .51666)
FT /tag= an
FT /product= "ORF#31 protein"
FT /note= "encoded protein shown in AAW72196"
FT complement(51701. .53575)
FT /tag= ao
FT /product= "ORF#32 protein"
FT /note= "encoded protein shown in AAW72197"
FT 54393. .58115
FT /tag= ap
FT /product= "ORF#33 protein"
FT /note= "encoded protein shown in AAW72198"
FT complement(58060. .58977)
FT /tag= aq
FT /product= "ORF#34 protein"
FT /note= "encoded protein shown in AAW72199"
FT complement(58970. .60760)
FT /tag= ar
FT /product= "ORF#35 protein"
FT /note= "encoded protein shown in AAW72200"
FT 60759. .61151
FT /tag= as
FT /product= "ORF#36 protein"
FT /note= "encoded protein shown in AAW72201"
FT 61241. .62071
FT /tag= at

Alignment Scores:

Pred. No.: 937 Length: 117213
 Score: 107.00 Matches: 62
 Percent Similarity: 32.00% Conservative: 18
 Best Local Similarity: 24.80% Mismatches: 93
 Query Match: 7.32% Indels: 77
 DB: 2 Gaps: 10

US-09-087-136-1 (1-275) x AAV62176 (1-117213)

```

QY      13 ProGluAspGlyAspArgAsnAlaArgGluAsnAspProLeuIleSerGlyGlyProLeu 32
DB      63985 CCGAGGGGTGGCGTCCGCGC-----CGGCACTCCCGCCCGCGTCCGCCGCGTTC 63932

QY      33 ProLeuGluSerProSerArgGlyLeuThrSerLeuLeuSerTyrAspProThrValPro 52
DB      63931 CCGCGGGTCTCTCCGACGACGCC-----CGCGCGGCTTGGGTTCTGCTTGGCCCTTCCGC 63902

QY      53 GluSerProAspMetLeuPheAlaArgGlyArgLeuGlyAsnLeuLeuThrThrIleLeys 72
DB      63901 CGGAGTCCG-----CGCCCCGGCTTGGGTTCTGCTTGGCCCTTCCGC 63857

QY      73 HisHisProSerGluIleIleGlyValLeuProGluAspTyrThrArgAlaAspGlu--- 91
DB      63856 CCGGCGCGCTCGCGTCCGCGGCGCTTCCCGCCCAAGCTCGCGCCCATCCGCGCTTC 63797

QY      92 GluPro-GlyArgGlnGlyArgProProGlyArgProArgGlyMetProArgHisGluSe 111
DB      63796 CGCGCCCGCAGAGAGAGCGTCCGCGCGGAGACACCGCGCGCGCGCGCCGCGCCAGACAA 63737

QY      111 rSerThrSerLeuMetGluSerProArgGlyThrMetThrAspSerLeysIleMetPh 131
DB      63736 CAGCGGCTCCCTCGCGACACCCCGCGGCGCGCGCGCGCCGACCCCGCGCGCTCA 63677

QY      131 eGluLeuArgGlyLysProPheGluMetIleAlaGlyArgPheGluGluGlyTyrSerLe 151
DB      63676 CCGCGCGCGCGGTCCG-----CTGCGCTCT 63653

QY      151 uGlyArgAlaTyrValLysGlyHisMetAsnAsnGluTyrGluProIleLys-AlaGlnA 171
DB      63652 CGGCTCGT-----GAACTCCCTCCCTCCCGCGGAGC 63617

QY      171 rGThrAspTyrAla-ProAsnLeuAlaValAspTyrLeuAlaCysArgGluIleHisArg 190
DB      63616 CAGCGGACACCGCGCGGCTTCCCGCGCAGCGCGGAGTTCGCGCTTCTCCGCGGCC 63557

QY      191 Met---ProArgProAspLysSerIleProGluLeuProIleValProSerArgIleAsp 209
DB      63556 TTGCCCCCGGACCTCCCGCGTCCGACGACG----- 63524

QY      210 GluPheAspAlaThrValAspProArgTyrGluThrAspLeuLysAsnGluTyrIleArg 229
DB      63523 -----CCCGCGCTT----- 63515

QY      230 HisTyrLysGlnValLys-----LysGlyTyrCysAlaHis 241
DB      63514 ---TGGCCCGACGCGCGGTGGCCCGGACGAGCCTTGTGCGGTTGGGTCTGCGCCGAG 63458

QY      242 GluArgArgTyrThrAlaProHis 249
DB      63457 GCCCGGTGGCGCGCGCGCCCGAC 63434

RESULT 12
AAD25519/c
ID      AAD25519 standard; DNA; 154746 BP.
XX      AAD25519;
XX      26-MAR-2002 (first entry)
XX      Human herpesvirus 2 complete DNA genome.
XX      Human herpesvirus 2, cytostatic; cancer; immunosuppressive; virucide;
KW
  
```

```

KW      antibacterial; fungicide; protozoacide; antirheumatic; antiinflammatory;
KW      antiarthritic; rheumatoid arthritis; neuroprotective; multiple sclerosis;
KW      immune response; vasotropic; vaccine; gene therapy; autoimmune disease;
KW      vasculitis; ds.
OS      Human herpesvirus 2.
XX      WO200176643-A1.
XX      18-OCT-2001.
XX      06-APR-2001; 2001WO-US011372.
XX      07-APR-2000; 2000US-0195680P.
XX      (BAYU) BAYLOR COLLEGE MEDICINE.
XX      Orson FM, Kinsey BM, Bhogal BS;
XX      WPI, 2002-066308/09.
XX      Composition for oral delivery of vaccines, comprises expression vector
XX      containing antigenic genomic sequence, bound to aggregated protein-
XX      polycationic polymer conjugate or suspension.
XX      Disclosure; Page 90-132; 145pp; English.
XX      The invention relates to a composition comprising an expression vector
XX      bound to an aggregated protein-polycationic polymer conjugate or
XX      suspension. The expression vector contains a promoter polynucleotide
XX      sequence operatively linked to a polynucleotide sequence encoding an
XX      antigen which is a fragment of a gene or genome associated with an
XX      infectious disease, cancer and autoimmune disease such as rheumatoid
XX      arthritis, vasculitis, and multiple sclerosis, pathogenic genomes
XX      consisting of bacterium, fungus, protozoa and virus such as human
XX      immunodeficiency virus (HIV), herpes simplex virus (HSV), hepatitis C
XX      virus (HCV), influenza and respiratory syncytial virus (RSV) and
XX      optionally comprising a nucleotide sequence encoding a cytokine (or a
XX      cytokine expression vector), is useful for inducing an immune response
XX      (systemic and/or mucosal) in an organism. The cytokine expression vector
XX      contains a sequence for granulocyte macrophage-colony stimulating factor
XX      (GM-CSF) or interleukin-12 (IL-12). The polynucleotide sequences encoding
XX      the antigen and the cytokine are under transcriptional control of same or
XX      different promoter polynucleotide sequences. The expression vector, as a
XX      DNA vaccine is useful for treating a condition in an organism. The
XX      present sequence is human herpesvirus 2 complete DNA genome related to
XX      the invention
XX      Sequence 154746 BP; 23003 A; 54218 C; 54701 G; 22824 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 1,32e+03 Length: 154746
Score: 107.00 Matches: 62
Percent Similarity: 32.00% Conservative: 18
Best Local Similarity: 24.80% Mismatches: 93
Query Match: 7.32% Indels: 77
DB: 6 Gaps: 10

US-09-087-136-1 (1-275) x AAD25519 (1-154746)
QY      13 ProGluAspGlyAspArgAsnAlaArgGluAsnAspProLeuIleSerGlyGlyProLeu 32
DB      72872 CCGAGGGGTGGCGTCCGCGC-----CGGCACTCCCGCCCGCGTCCGCCGCGTTC 72819

QY      33 ProLeuGluSerProSerArgGlyLeuThrSerLeuLeuSerTyrAspProThrValPro 52
DB      72818 CCGCGGGTCTCTCCGACGACGCC-----CGCGCGGCTTGGGTTCTGCTTGGCCCTTCCGC 72789

QY      53 GluSerProAspMetLeuPheAlaArgGlyArgLeuGlyAsnLeuLeuThrThrIleLeys 72
DB      72788 CGGAGTCCG-----CGCCCCGGCTTGGGTTCTGCTTGGCCCTTCCGC 72744

QY      73 HisHisProSerGluIleIleGlyValLeuProGluAspTyrThrArgAlaAspGlu--- 91
  
```

Db	72743	CGGACCCCGCTCGGCTCGCGGAGCCCTTTCGCCGCCACGCTCGCGCCCATCCGCGCTC	72684
Qy	92	GIUPro-GIYAArgInGIYAArgProProGIYAArgProAArgLYMeProAArgHIEglUSe	111
Db	72683	CGCCCGCCGAAAGAGACGGCTGCTCTCCCGGAGGCAACCGCCGCGGCGCCCGCCAGACA	72624
Qy	111	rSErThrSerLeuMetGluSerProAArgLYrThrMetThrArgAAspSerLYeIleMerp	131
Db	72623	CAGCGGGTCCCTCGCCACACCCCGCGGAGCCCGCGGCGGAGCCACCCCGGCGCTCA	72564
Qy	131	egIUleuAArgLYuSerProPhegluMetIleAglYAArgPhegluIugluIuTrSerIe	151
Db	72563	CGCGCCCGCGGCTGC-----CTCGCTCT	72540
Qy	151	uGIYAArgAlaTrpValLYeGlyHISMetAAsnGluTrYGluProIleLYe-AlaGlnA	171
Db	72539	CGGCTCGGT-----GAATCTCCCTCCCTCTCCCGGAGAC	72504
Qy	171	rGThrAAspTYrAla-ProAAsnLeuAlaValAAspTYrLeuAlaCYsAArgGluIleHISArg	190
Db	72503	CAGCGGACACACGCGCGCGGTTTCGCGCGACGCGGCGGAGHCGCGCTTCGCGGCG	72444
Qy	191	Met---ProAArgProAAspLYSerIleProGluLeuProIleValProSerArgIleAsp	209
Db	72443	TTGGCCCCCGACCTCGCGCTCCAGACCGCC-----	72411
Qy	210	GIUleuAAspAlaThrValAAspProAArgTYrGluThrAAspLeuLYsAsnGluTrIleArg	229
Db	72410	-----CCCCGCTT-----	72402
Qy	230	HIaTrpLYsGlnValLYs-----LYsGlyTYrCYsAlaHIS	241
Db	72401	---TGGCCCCAGCGCGGTTGCCCGGAGCGAGCGCTTGTGGCGGTGGTGGTGGCGGAG	72345
Qy	242	GIuAArgAArgAArgThrAlaProHIS	249
Db	72344	GCCCGGTTGCCCGCGGCCCGCCAC	72321
RESULT 13			
ABD17925/c			
ID	ABD17925	standard; DNA; 1908 BP.	
XX	XX	ABD17925;	
XX	AC		
DT	29-JUL-2004	(first entry)	
XX	XX		
DE	Pseudomonas aeruginosa polymnucleotide #16529.		
XX	XX	Bacterial infection; gene; de; Pseudomonas aeruginosa infection;	
KW	antibacterial.		
XX	XX		
OS	Pseudomonas aeruginosa.		
XX	XX		
FN	US6551795-B1.		
XX	XX		
PD	22-APR-2003.		
XX	XX		
PF	18-FEB-1999;	99US-0025291.	
XX	XX		
PR	18-FEB-1998;	98US-0074788P.	
XX	XX		
PR	27-JUL-1998;	98US-0094190P.	
XX	XX		
PA	(GENO-) GENOME THERAPEUTICS CORP.		
XX	XX		
PI	Rubenfield WU, NoIling J, Deloughery C, Bush D;		
XX	XX		
DR	WPI; 2003-615309/58.		
XX	XX		
DR	P-PSDB; ABO84354.		
XX	XX		
PT	Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide, useful as molecular targets for diagnostics, prophylaxis and treatment of pathological conditions resulting from bacterial infection.		

XX Disclosure; SEQ ID NO 16529; 455bp; English.

XX The invention relates to *Pseudomonas aeruginosa* polypeptides and the

CC polynucleotides encoding them. The sequences are useful in diagnosis and

CC therapy of pathological conditions, as molecular targets for diagnostics,

CC prophylaxis and treatment of pathological conditions resulting from a

CC bacterial infection, for evaluating a compound, such as a polypeptide, for

CC the ability to bind a P. aeruginosa nucleic acid, as components of

CC effective antibacterial targets, as targets for antibacterial drugs,

CC including anti-P. aeruginosa drugs, as templates for recombinant

CC production of P. aeruginosa-derived peptides or polypeptides, as target

CC components for diagnosis and/or treatment of P. aeruginosa-caused

CC infection, and in detection of P. aeruginosa sequences or other sequences

CC of *Pseudomonas* species using biochip technology. Sequences ABD01397-

CC ABD1967 represent P. aeruginosa polynucleotides of the invention. Note:

CC The sequence data for this patent did not form part of the printed

CC specification but was obtained in electronic format from USPTO at

CC seqdata.uspto.gov/sequence.html

XX

XX

Seq Sequence 1908 BP; 295 A; 622 C; 700 G; 289 T; 0 U; 2 Other;

Alignment Scores:

Prod. No.: 6.18 Length: 1908

Score: 106.50 Matches: 64

Percent Similarity: 33.45% Conservative: 30

Best Local Similarity: 22.78% Mismatches: 86

Query Match: 7.28% Indels: 101

DB: 11 Gaps: 16

US-09-087-136-1 (1-275) x ABD17925 (1-1908)

QY 27 IleserGIyGIyPro-----LeuProLeuGluSerProSerArgLys 40

DB 1894 GTTTCGGAGGAGACCCCGCCAGACCAACTACTCTTCACTCCCAATCGGCATCGCGCGCA 1835

QY 41 LeuthSerLeuLeuSerTyrAspProThyValProGluSerProAbpMetLysPheAla 60

DB 1834 -----AGTGTCTGGCACCAACCTCGCTCGTGAAGC----- 1808

QY 61 ArgLysArgLysLeu-----GlyAanLeuLeuThrThr 70

DB 1807 CGTCCGCCGGAATTCAGCAAGCACTGCGCTGCTGCTGTGTGGCGGCTTACGTCTCGC 1748

QY 71 IleLysHis-----HisProSerGluIleIleGlyValLeuProGlu 84

DB 1747 CTTACGACATCTCGTCTGCTGCTGATCGACACCTCCCGTGGCGCTGTCTTA-----CGC 1694

QY 85 AspTyrThrArgAlaAspGluGluProGlyArgGlnGly-----ArgPro 99

DB 1693 CATCTGCTCCGGAATCGGACATCTGCTGTGTGACCTGTGGCGATGTTCTTCATCGGCCA 1634

QY 100 ProGlyArgProAlaGlyLysMetProArgHisGluSerSerThrSerLeuMetGluSerPro 119

DB 1633 GCGCCTGGACCCCGCCGCTCTCTCGGATCGGCTGATCATCGCCGCGCTGTGTGAT 1574

QY 120 ArgLysThrMetCysArgAspSerLysIleMetPheGluLeuArgGlyLys----- 136

DB 1573 CCAAGTTGTTCTCCCGCGCTTGGGGACATGAAC---GACCTCCGCGGCGGAGTTGTAT 1517

QY 137 -----PropheGluMetIleAlaGlyArgPheGluGluGluLysSerLeuGlyArgAla 154

DB 1516 ACTGGCCCTTTCGCTTC----- 1499

QY 155 TrpValLysGlyHisMetAsnAsnGluLysGluProIleLysAlaGlnArgThyAspTyr 174

DB 1498 -----CCGATA-----AGCAGACCCCAT 1481

QY 175 AlaProAenLeuAlaValAspTyrLeuAlaCysArgGluIleHisArgMetProArgPro 194

DB 1480 GCCCCNAGC-----CTTTCCACCGACATCTCTATCTGTGGCGGCGGTAT 1436

QY 195 AspLysSerIleProGluLeuProIleValProSerArgIle---AspGluPheAspAla 213


```

Dd      955 GGCCTTACCTCCGCCCA-----CGACTATGGCGACCGAGCCGATGAA 999
Qy      267 ThrValGlnGlnAlaLeuGly 273
          ::::
Dd      1000 CGACTTCGGCGACCCCGCTGGT 1020

RESULT 15
ACN44556/C
ID      ACN44556 standard; DNA; 22262 BP.
XX      ACN44556;
XX      ACN44556;
XX      18-NOV-2004 (first entry)
XX      Mouse genomic sequence MCG21411.
XX      Cytosarctic; carcinoma; lymphoma; cancer; murine; gene; ss.
XX      Mus musculus.
XX      WO2003073826-A2.
XX      12-SEP-2003.
XX      28-FEB-2003; 2003WO-US006235.
XX      01-MAR-2002; 2002US-00087192.
XX      (SAGR-) SAGRES DISCOVERY.
XX      Morrie DW;
XX      WPI; 2003-328604/31.
XX      Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT      comprises a nucleotide sequence.
XX      Claim 1; SEQ ID NO 1063; 0pp; English.
XX      The present invention relates to novel DNA and protein sequences which
CC      are associated with carcinomas. The sequences are useful for: (i) for
CC      screening drug candidates; (ii) for screening of bioactive agent capable
CC      of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC      a bioactive agent capable of modulating the activity of CAP; (iv) for
CC      evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC      carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC      carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a Diolip;
CC      (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC      determining Carcinoma Associated (CA) gene copy number. In addition, the
CC      CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC      carcinoma including lymphoma. The present sequence is one such CA coding
CC      sequence. Note: This patent is an equivalent to basic patent
CC      US2002182586A1, for which no sequence data was published
XX      SQ
XX      Sequence 22262 BP; 5720 A; 5332 C; 5185 G; 5985 T; 0 U; 40 Other;

Alignment Scores:
Pred. No.:      181      Length:      22262
Score:          105.00   Matches:      65
Percent Similarity: 35.27% Conservative: 26
Best Local Similarity: 25.19% Mismatches: 89
Query Match:      7.18% Indels:      78
DB:              11      Gaps:      15

US-09-087-136-1 (1-275) x ACN44556 (1-22262)
Qy      28 SerGlnGly-----ProLeuProLeuGlnSerProSerArgGlyLeu 41
          |||||
Dd      10107 AGTGGGGGAGTCGAGGAGGAGGTGTCCTTACCT-----CCCGCGAGAGCGTA 10057
          |||||
Qy      42 ThSerLeuLeuSerTyrAspProIleValProGlnSerProAspMetIleYsPheAlaG 61
          |||||

```

```

Db      10056 ACGGTGGGGCTTTCT-----CTGGGG 10036
QY      62  LyaArgLeuGIaYbAnLeuLeuThrThrIleYbNHsHPProSerGIuIleIleGIaY  81
Db      10035 CGAGCA-----GCCAACAACCTTCTCCGGTATACCTTCAAGATTGGCCAAAGTT  9988
QY      82  LeuProGIuApyrThrThrArgIaAaerGIuGIuAProGIaYArgIuGIuYArgProPro----  100
Db      9987 TCTCGAGCCCCGGGGCGACGCCATACGACAGTCCCTTAACCGACGACGAGCTCCGCCA  9928
QY      101  --GIaYrProArgIuYbMetProArgNHs-----GluSer  112
Db      9927 GGCACGAGCCCTCGCTGCTCTGCTGCTCGCCGCGCGCTGCGCTGCTGCTGCTGCTCGG  9868
QY      113  ThrSerIuMetGluSerProArgIuYbThr-----MetThrArgAspSerIuIleMet  130
Db      9867 CTTTCTCTCTTGAAGCCCCCGCGAGCGACATGCAAGCGCGCTCTGTATTTCTGAGCTCC  9808
QY      131  PheGIuLeuAArg--GIaYsPro-----PheGIuMetIleAaGIaYArgPheGIu  146
Db      9807 TTCAAGCTCTAAGAAAGTCGGCGCCCAAGTTCAAGTCGAATAAGCTGCTTCTCGCTTTGAAT  9748
QY      147  --GIuGIuYrSerLeuGIaYArgAlaTPrValIySeGIuNHsMetAaAenGIuYrGIuP  166
Db      9747 CTGATCGGCTGCTCGGAGGTAAAGCTGACCAAGCTGTTTGGAAACGAGAT-----  9695
QY      166  roIleYbAlaGIaArgThrAspYrAlaProAenLeuAValaApyrYrLeuAlaCysA  186
Db      9694 -----TGCACGGAATCCGAAGCCCGA  9673
QY      186  rGtGIuIleNHsArgMetProArgProApyrSerIleProGIuLeuPro---IleValP  205
Db      9672 GGCCTCTGCAC-----CCGCGGCTCTGCTGACCGCGGCTCTCTCGCCCGCGAGTCC  9619
QY      205  roSerArgIleAaerGIuPheAAspAlaThrValAspProAArgYrGIuThrApyrLeuYbA  225
Db      9618 CAGCCGCC-----GACTCTCGCGCTCCACACCGCCCGCGCAAGGAAGACCGCGAGTA  9565
QY      225  snGIuYrIleArgNHsIleTPrYbGIuValIySeGIuYr-----  238
Db      9564 ACGGCTGGAGACCGCACTGG-----TGGCTAAGAAAGAGCTGGCGCT  9523
QY      239  -----CysAlaNHsGIaYArgThrThrAlaProNHsAlaAysSer  252
Db      9522 TGCAGGAGCTCGGAGCGCGCCGCGCCCTCCCTCCCTCCCTCCCTTACCAACAC  9473

```

Search completed: April 9, 2005, 08:18:57
Job time : 664 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus2 model

Run on: April 9, 2005, 07:57:27 | Search time 4654 seconds

(without alignments)
2863.168 Million cell updates/sec

Title: US-09-087-136-1

Perfect score: 1462
Sequence: 1 MSEDPLAEFLPEDGDNRN.....INKIYQGESKTVQALGLI 275

Scoring table:

BIOSUM62
Xgapop 10.0, Ygapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=xlh
-Q=/cgn2.1/USPRO.spool/US09087136/runat_07042005_163616_903/app_query.fasta_1.455
-DB=genemb1 -QMT=faastap -SUFFIX=rge -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biolum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTMT=ptco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09087136@cgn.1_14200@runat_07042005_163616_903 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmb1:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_on:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
C 1	1225.5	83.8	32981	3 U00047	U00047 Caenorhabdi
C 2	124.5	8.5	6339	6 CQ722208	CQ722208 Sequence
C 3	117	8.0	196857	5 AL928885	AL928885 Zebrafish
C 4	116	7.9	42074	1 SGR578458	AD578458 Streptomy

C 5	115	7.9	240105	5 BX571714	BX571714 Zebrafish
C 6	113	7.7	181667	2 AC146961	AC146961 Otolomur
C 7	112	7.7	11031	1 AE004648	AE004648 Pseudomon
C 8	111.5	7.6	1045	8 BT005099	BT005099 Arabidops
C 9	111.5	7.6	1324	8 AK117550	AK117550 Arabidops
C 10	111	7.6	109446	10 AL645845	AL645845 Mouse DNA
C 11	111	7.6	199326	10 AC005528	AC005528 Mus muscu
C 12	110.5	7.6	36270	1 SLINC	SLINC
C 13	110.5	7.6	41987	1 AY498874	AY498874 Streptomy
C 14	110.5	7.6	69301	1 STU82965	STU82965 Streptomyce
C 15	110.5	7.6	110000	1 AE017282_24	AE017282_24
C 16	109.5	7.5	1639	6 BT009343	BT009343 Sequence
C 17	109.5	7.5	1707	6 AR195051	AR195051 Sequence
C 18	109	7.5	1234	8 AY099812	AY099812 Arabidops
C 19	109	7.5	1318	8 AY085490	AY085490 Arabidops
C 20	109	7.5	166136	2 AC140964	AC140964 Pan trogl
C 21	109	7.5	181054	2 AC130189	AC130189 Pan trogl
C 22	108.5	7.4	1778	8 ZMCAT2R	ZMCAT2R
C 23	108	7.4	3483	2 AC017763	AC017763 Drosophill
C 24	107.5	7.4	6906	6 AX344602	AX344602 Sequence
C 25	107.5	7.4	15397	1 AB093554	AB093554 Streptomy
C 26	107.5	7.4	299800	1 AP005028	AP005028 Streptomy
C 27	107	7.3	5135	1 SERDNABP	SERDNABP
C 28	107	7.3	154746	14 HSV2HG52	HSV2HG52
C 29	107	7.3	205794	9 AC010864	AC010864 Homo sapi
C 30	106.5	7.3	10459	1 AE004717	AE004717 Pseudomon
C 31	106.5	7.3	12046	1 AE005017	AE005017 Halobacte
C 32	106.5	7.3	110000	1 BX571965_19	BX571965_19
C 33	106	7.3	109192	8 CNS08CB7	CNS08CB7
C 34	106	7.3	180599	8 CNS08CB7	CNS08CB7
C 35	106	7.3	298550	1 AP005029	AP005029 Streptomy
C 36	106	7.3	299925	1 AP005048	AP005048 Streptomy
C 37	105.5	7.2	183155	2 AL591169	AL591169 Homo sapi
C 38	105.5	7.2	205973	10 AL772255	AL772255 Mouse DNA
C 39	105.5	7.2	299925	1 AP005042	AP005042 Streptomy
C 40	105	7.2	3132	10 BC083806	BC083806 Rattus no
C 41	105	7.2	146453	4 AY152827	AY152827 Felis cat
C 42	105	7.2	176113	2 AC132574	AC132574 Mus muscu
C 43	104.5	7.1	2888	6 AA8323	AA8323 Sequence 1
C 44	104.5	7.1	2888	6 AR198353	AR198353 Sequence
C 45	104.5	7.1	13901	1 AY258009	AY258009 Streptomy

ALIGNMENTS

RESULT 1	U00047	32981 bp	DNA	linear	INV 11-SEP-2004
LOCUS	U00047/c				
DEFINITION	Caenorhabditis elegans cosmid ZK418, complete sequence.				
ACCESSION	U00047				
VERSION	U00047.1	GI:470368			
KEYWORDS	HTG.				
SOURCE	Caenorhabditis elegans				
ORGANISM	Caenorhabditis elegans				
REFERENCE	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;				
AUTHORS	Rhabditoidae; Rhabditidae; Pelodierinae; Caenorhabditis.				
CONSRTM	1 (bases 1 to 32981)				
TITLE	WormBase Consortium				
REFERENCE	Genome sequence of the nematode C. elegans: a platform for				
AUTHORS	investigating biology. The C. elegans Sequencing Consortium				
TITLE	Science 282 (5396), 2012-2018 (1998)				
REFERENCE	JOURNLT MEDLINE PUBMED				
AUTHORS	99069613				
TITLE	9851916				
REFERENCE	2 (bases 1 to 32981)				
AUTHORS	Fulton, L.				
TITLE	The sequence of C. elegans cosmid ZK418				
REFERENCE	Unpublished (2001)				
AUTHORS	3 (bases 1 to 32981)				
TITLE	Waterston, R.				
REFERENCE	Submitted (08-APR-1994) Department of Genetics, Washington				
JOURNAL	University, Genome Sequencing Center, 4444 Forest Park Avenue, St.				

REFERENCE AUTHORS TITLE JOURNAL	Louis, MO 63110, USA 4 (bases 1 to 32981) Waterston,R. Direct Submission Submitted (28-JUN-2001) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
REFERENCE AUTHORS TITLE JOURNAL	5 (bases 1 to 32981) Waterston,R. Direct Submission Submitted (29-MAY-2002) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
REFERENCE AUTHORS TITLE JOURNAL	6 (bases 1 to 32981) Waterston,R. Direct Submission Submitted (22-NOV-2002) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
REFERENCE AUTHORS TITLE JOURNAL	7 (bases 1 to 32981) Wilson,R. Direct Submission Submitted (15-JUN-2003) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
REFERENCE AUTHORS CONSRITM TITLE JOURNAL	8 (bases 1 to 32981) WormBase Consortium Direct Submission Submitted (11-SEP-2004) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
COMMENT	Submitted by: Genome Sequencing Center Department of Genetics, Washington University St. Louis, MO 63110, USA, and Sanger Centre, Hinxton Hall Cambridge CB10 1RO, England email: submissions@watson.wustl.edu and jee@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

For a graphical representation of this clone sequence and its analysis see:
<http://www.wormbase.org/db/seq/sequence?name=ZK418;class=Sequence>

NEIGHBORING CLONE INFORMATION

The 5' clone is R01H2, 200 bp overlap; the 3' clone is B0280, 200 bp overlap. Actual start of this clone is at base position 197 of ZK418; actual end is at 7055 of B0280.

NOTES:

Coding sequences below are the result of integration and manual review of the following data: computer analysis using the program GeneFinder (P. Green and L. Hillier, personal communication), the large scale BSP projects of Yuji Kohara (http://www.ddb.jri.ac.jp/c-elegans/html/CE_INDEX.html) and the C. elegans ORFome cloning project (<http://worldb.dicr.harvard.edu/>), similarity to other proteins from Blastx analyses (<http://blast.wustl.edu/>), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research

10:1115-1125, 2000), individual C. elegans GenBank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SB (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).
Location/Qualifiers
1..32981
/organism="Caenorhabditis elegans"
/mol_type="genomic DNA"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="III"
/clone="ZK418"
92..1634
/gene="ZK418.3"
/locus_tag="ZK418.3"
/join(92..178,1121..1234,1283..1405,1473..1634)
/gene="ZK418.3"
/locus_tag="ZK418.3"
/standard_name="ZK418.3"
/note="contains similarity to Mus musculus Transmembrane protein 17.; TR:08X0U3; coded for by the following C. elegans cDNAs: YK868e01.3"
/codon_start=1
/product="Hypothetical protein ZK418.3"
/protein_id="AAA50687.1"
/db_xref="GI:470371"
/db_xref="WormBase:ZK418.3"
/translation="MADFSCINVCIPCNSCSLRLQNGCCPRHLSLSLTYQVAFSA CLVHLGSGVGLGFGYGNLAENNSALFGFLITITIIQIPLTVLANGSFNLPLEY VFYMTVFAIEMAKTAREEVAKYINKIOEKVADLEKRTSGKSPKSGRTKKQLS R"
complement(1751..3301)
/gene="lin-37"
/locus_tag="ZK418.4"
complement(join(1751..1868,2059..2228,2276..2555,2993..3127,3177..3301))
/gene="lin-37"
/locus_tag="ZK418.4"
/standard_name="ZK418.4"
/note="contains similarity to Arabidopsis thaliana Hypothetical protein Atg32840.; TR:08GYK6; coded for by the following C. elegans cDNAs: CB102055, CB102059, CB102062, CB102064, CB102065, CB102067, CB102169, CB102170, CB102172, CB102173, CB102175, CB102177, OSTF038D2_1, YK71d12.5, YK90a5.5, YK96b5.3, YK96b5.5, YK19116.3, YK550h7.3, YK550h7.5, YK560h10.3, YK560h10.5, YK868e01.3"
/codon_start=1
/product="Abnormal cell lineage protein 37"
/protein_id="VAA50688.2"
/db_xref="GI:14574553"
/db_xref="WormBase:ZK418.4"
/translation="MSRIDPLAEFLLEDGDRARQNDPLISGGPLPLEBSPKRLTSLSYPTVPEBSPDMFKARKLGNLTITKHPSLITGYLPDYGRADDEPGKGRPKRPMRHSSTLMSRPMTRDSKIMPELGRPFEMIGFEESYSLGAWYKGMNNEPPIKAQRTDYPYLAVDYLAICREIHMPRDSKIPELPIVPSIDEDATVDPRYETDLKNEYIRHMKQVKKGCACQRRTAHSALINIKTYGSESKTYEQALGL"
complement(3301..3302)
/note="SL2 trans-splice acceptor; see YK90a5.5"
complement(3582..4582)
/gene="ZK418.5"
/locus_tag="ZK418.5"
complement(join(3582..3717,3768..3904,3955..4152,4330..4463,4510..4582))
/gene="ZK418.5"
/locus_tag="ZK418.5"
/standard_name="ZK418.5"
/note="contains similarity to Mus musculus Mus musculus adult male small intestine cDNA, RIKEN full-lengthenriched library, clone:2010004E11
product:SEVEN TRANSMEMBRANE DOMAINPROTEIN homolog.; TR:09B0F0; coded for by the following C. elegans cDNAs: OSTF037F8_1, OSTR037F8_1, YK329h10.3, YK329h10.5,

FEATURES
source
gene
CDS
misc_feature
gene
CDS

Yk600e3.3, Yk600e3.5"
 /codon_start=1
 /product="Hypothetical protein ZK418.5"
 /protein_id="AA50689.2"
 /db_xref="GI:14574554"
 /db_xref="Molbase:ZK418.5"
 /translation="MSFPHINCPALAPAFYIVYKSGINESSIMKCATASGYLL
 TOLAKLIATPEPPALDSEGPSIVPEPKSSADIIDVGLHLMNPLAGGEVYV
 GIGWGPASVVAHRLVILMGARGATFWRWQVSTLSDSADLVIVSLACTMTTTRT
 PNKLIVSPILAMCVSFIVYQVQHTFSLYGLMSLAFPAISIAITAILTVVYSANRT
 ASTRKNE"
 misc_feature
 complement (4600, .4601)
 /note="SL1 trans-eplice leader (TTC-RED); see
 cel031703ab0112"
 complement (5033, .9583)
 /gene="ZK418.6"
 /locus_tag="ZK418.6"
 complement (5033, .5578, 5882, .6128, 6423, .6777,
 6943, .7091, 7366, .7533, 7582, .7789, 8102, .8318, 8371, .8501,
 8555, .8718, 8764, .8897, 9314, .9400, 9512, .9583)
 /gene="ZK418.6"
 /locus_tag="ZK418.6"
 /standard_name="ZK418.6"
 /note="contains similarity to Pfam domain PF00001 (7
 transmembrane receptor (rhodopsin family)); coded for by
 the following C. elegans cDNAs: OSTR050H5_1"
 /codon_start=1
 /product="Hypothetical protein ZK418.6"

Alignment Scores:

Pred. No.: 4,986-77 Length: 32981
 Score: 1225.50 Matches: 273
 Percent Similarity: 52.70% Conservative: 0
 Best Local Similarity: 52.70% Mismatches: 2
 Query Match: 83.82% Indels: 245
 Gaps: 4

US-09-087-136-1 (1-275) x U00047 (1-32981)

QY 1 MetSerGluIleAspProLeuAlaGluPheLeuLeuProGluAspGlyAspArgAsnAla 20
 Db 3301 ATGTGAGAAATAGATCCACTGCGAGTCTTGCTTCAGAAAGACGAGATCGAATGCT 3242
 QY 21 ArgGlnAsnAspProLeuIleSerGlyGlyProLeuProLeuGluSerProSerArgLys 40
 Db 3241 CGTCAAATGATCCATGATTAAGCGGAGGTCACCTTCATTGGAATCGCCAGACAGAA 3182
 QY 41 LeuThr-----SerLe 44
 Db 3181 CTCACGTAATCTTCTAAACTGGAACAATTGCTACTCAAGATGCTCCAGATCCCT 3122
 QY 44 uLeuSerTyAspProThrValProGluSerProAspMetIleAspGlyValGlyLe 64
 Db 3121 GTTATCCATGATCCGACAGATCCGAGATGACCGATATGAAATTCGCGCAAAAACGTCT 3062
 QY 64 uGlyAsnLeuLeuThrIleValHisIleProSerGluIleIleGlyValLeuProG 84
 Db 3061 GGGAAATCTGCGAACACATTAACATCACCATCGAAATTAATTTGAGTACTCCAGA 3002
 QY 84 uAspTyThr-----TPC 239
 Db 3001 AGATTATAC-GTAACTGATAAATGGAATTTCAATTCTTGTTAATTTTCGCGAGAGT 2943
 QY 87 ----- 87
 Db 2942 ATACCATGCTGGGCGCAAAATCATTTTTTTGTTTGATTGATTGTGATCAAAAGCA 2883
 QY 87 ----- 87
 Db 2882 AAAATATGTAAAAAGCCAAAAAAATTGAGATTTGAATTTTAAACCTTAATATTGAG 2823
 QY 87 ----- 87

Db 2822 AAAAAGATTAAAACTATGACAGACACATTAGTTGAAAGAAAAAGTAAATTAATTT 2763
 QY 87 ----- 87
 Db 2762 TGTACACAGAAAACGTTTACAGTAATAACAGCATTTTTTTTCAGTTTTTGTTCAGAT 2703
 QY 87 ----- 87
 Db 2702 AAATAGTTAATTTTGTCAAATGTCCGCTCTCGAATATACATAACAGAAAAAATTTAG 2643
 QY 87 ----- 87
 Db 2642 ACAATGACACCTTTTGTGTTTTCGTAATTAATTTTACACATTAATGAAAAATTAGAG 2583
 QY 88 -----ArgAlaAspGluGluProGlyValGlyGlnGlyArg 98
 Db 2582 TTTCCATTGCATTTTTTAAATTTTCAGTCGTGCGATTAAGAGCCCGGCGCAAGAGCG 2523
 QY 98 GProProGlyArgProArgLysMetProArgHisGluSerSerThrSerLeuMetGlyLe 118
 Db 2522 TTCACACAGTGGCCCTCGTAAGATGCCCGGTACAGAACTTTCACCTTACCTTATGGAATC 2463
 QY 118 rProArgLysThrMetThrArgAspSerLysIleMetPheGluLeuArgGlyLysProPh 138
 Db 2462 ACCACGCAAGACTATGACTCGGATTTAAATTTGTTGAATTCGCGTGAAGAAACATTT 2403
 QY 138 eGluMetIleAlaGlyArgPheGluGluGluIleValSerLeuGlyValGlyAlaTPValLysG 158
 Db 2402 CGAATGTACTGACCTGACCTTTTGAAGAAATATTCACCTTGTTAGACATGCTTTAAAG 2343
 QY 158 YHisMetAsnAsnGluTyArgIleProIleValAlaGlnArgThrAspTyArgAlaProAsnLe 178
 Db 2342 ACACATGATATATGATATGAAACATTAAGCTTAAGGACAGCATTTGACCCAGATCT 2283
 QY 178 uAlaVal-----AspTy 182
 Db 2282 GGCTGT-TGTGAGTGGGAAAGATTAATGAAAAATGTTAATTAATTTCTTCAGAGATTA 2224
 QY 182 rLeuAlaLysArgGluIleHisArgMetProArgProAspLysSerIleProGluLeuArg 202
 Db 2223 TCTTCATGATGCGCGAGATTCAGATCGAATGCAACGCTCCAAATTAATTTCCGACACTGCC 2164
 QY 202 oIleValProSerArgIleAspGluPheAspAlaThrValAspProArgTyArgLysThrAs 222
 Db 2163 AATTGTTCACTTGAATTCAGATGAATTCGACGCTTCACTGATCCAAAGTATGAACACA 2104
 QY 222 pLeuLysAsnGluTyArgIleArgHisTPValGlnVal-LysLysGly----- 237
 Db 2103 TTTGAAAAATGATATCATTCGTCATGGAACAAAGTCAAAAAAGGTAAACACTTCTTCTG 2044
 QY 237 ----- 237
 Db 2043 CCATTGAAAAATTCGTAATAATTTATTTCTTAAAGTATTAACGTAACTGTGCTATAGT 1984
 QY 237 ----- 237
 Db 1983 TAGAATTTCAATTTGTCGATTTTGTATCAAACTAATAAATAATTTGTTACTAGCA 1924
 QY 238 -----TPC 239
 Db 1923 TCTGTGATCAGTACTATTCTTCAAAAGTCAGAGATCCAAAAAATAATTTCAAGTTGGT 1864
 QY 239 yValAlaHisGlnArgArgArgThrAlaProHisAlaArgSerIleAlaLeuIleAsnLys 259
 Db 1863 GTGCTCATCAACGTCGTCGACTGCTCCCAATGCAAGAAAGCATATGATTAATCAACAAA 1804
 QY 259 lTyArgLysProGlyGluSerLysThrValGluGlnAlaLeuGlyLeuLe 275
 Db 1803 TCTACACGCTGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1754
 RESULT 2
 CQ722208
 LOCUS CQ722208 6339 bp DNA linear PAT 03-FEB-2004

```

DEFINITION Sequence 8142 from Patent WO02068579.
ACCESSION   C0722208
VERSION     C0722208.1
KEYWORDS    GI:42283065
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1
AUTHORS    Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
TITLE      Kits, such as nucleic acid arrays, comprising a majority of
            humanexons or transcripts, for detecting expression and other uses
            thereof
JOURNAL    Patent: WO 02068579-A 8142 06-SEP-2002;
            PE Corporation (NY) (US)
FEATURES
  source     1. 6339
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.:      35, 6
Score:          124.50
Percent Similarity: 40.60%
Best Local Similarity: 25.17%
Query Match:    8.52%
DB:             6
            Gaps: 18

US-09-087-136-1 (1-275) x C0722208 (1-6339)
QY      1 MetSerGluIleAspProLeuAlaGluPheLeuProGluAspGlyAspArgAsn--- 19
      Db 2143 CTGAAGCAGCTGTAAGAGCAGGAGGCTGCGCATCCGCCCATGAGCAGTACCGCTCTCCG 2202
      QY 20 ---AlaArgGluAsnAspProLeuIleSerGlyGlyProLeuPro-----LeuGluSer 36
      Db 2203 GCCCCACATGGAACACATGTCCTCCACAGGTCACCCCTCCCTGGGCTTCATGAC 2262
      QY 37 ProSerArgLysLeuThrSerLeuLeuSerTyrAspProThrValProGluSerProAsp 56
      Db 2263 ACTGCCCGG-----CCGTCCTCGCGCGGCCCAACCCATCTCCAAACCCGCTCC 2313
      QY 57 MetLysPheAlaArgLysArgLeuGlyAsnLeuLeuThrIleLysHisProSer 76
      Db 2314 -----IleGlyValLeu-----CTCATCTCCTCGCAAG---CACCCACGC 2340
      QY 77 GGuile-----IleGlyValLeu-----ProGluAsp----- 85
      Db 2341 GTCTCGAGAGGCAAAATAGTCCATCTCCCAAGGACGACGCTGATCTCTGTACAG 2400
      QY 86 -----TyrThrArg-----AlaAspGluGluProGlyArgGlnGly---ArgProPro 100
      Db 2401 GGCACCATCACCAGGATCATCGGCGAGGACAGCCCGGAGTCTT-GGACCGCGGCGGGA 2459
      QY 101 GLYAGProArgLysLysMetProArgHis-----GluSerSerThrSerLeu--- 115
      Db 2460 GACAGCCTGCGCCAAAGGCGACGTCATCTACGAAGGCAAGAGG-GCCACGCTTGTCT 2518
      QY 116 -----MetGluSerProArgLysThr-----MetThrArgAsp 126
      Db 2519 ATGAGGCGGCGCATGCTGTGACCCAGTCTCTCCAGAGAGGAGCGAGCAAGCAGCTCAGGAC 2578
      QY 127 SerLysIleMetPheGluLeuArgGlyLysArgProPheGluMetIleAlaGlyArgPhe--- 145
      Db 2579 CCCCCCATGAGAGC-GCCGCCCCCAAGGCGACATATGACATGATGAGGGCGCGGTGGC 267
      QY 146 -----GluGluGluTyrSerLeuGly 152
      Db 2638 AGAGCCATCTCTCAGCAGCAGCATGGAAGGTATGACAGGAGGAGACTTCATCTCTCGGT 2697
      QY 153 ArgAlaTrpValLysGlyHisMetAsnAsnGluTyrGluProIleLysAlaGlnArgThr 172

```

```

      Db 2698 GCCCCTGTGGCGCGCTACAGAGCAGCGCCCTGGCGCCCTGTAAGCTG---AAGCG 2754
      QY 173 AspTyrAlaProAsnLeuAlaValAspTyrLeuAlaCysArgGluIleHisArgMetPro 192
      Db 2755 GCCCATGAGGCGCTGTGTGCGCACCGGTGAAGAGGCGGCCCTCCATCTACGAAAGATCCCG 2814
      QY 193 ArgProAsp---LysSerIleProGluLeuProIleValProSerArgIleAspGluPhe 211
      Db 2815 CCGGAGGAGCTGCGCGACACCGCCGAGCTGCGCCCGCGCGCGCGCTCAAGAG--- 2871
      QY 212 AspAlaThrValAspProArgTyrGluThrAspLeuLysAsnGluTyrIleArgHisTrp 231
      Db 2872 ---GGCTCATTCAGCAGCGG-CACCCGCTCAAGTATGACACCGCGCGCTCACCTGG 2927
      QY 232 LysGluValLysLysGlyTyrCysAlaHisGlnArgArgArgThrAlaProHis 249
      Db 2928 CTCGA-----AAACACAGACGTAAGCTCTCTCAT 2957

RESULT 3
AL928885/c
LOCUS      AL928885
DEFINITION Zebrafish DNA sequence from clone CH211-195G24 in linkage group 3,
            complete sequence.
ACCESSION   AL928885
VERSION     AL928885.9
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
ORGANISM    Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
            1 (bases 1 to 196857)
REFERENCE   1
AUTHORS    Cordy, N.
TITLE      Direct Submission
JOURNAL    Submitted (06-AUG-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
            On Aug 6, 2003 this sequence version replaced gi:2789631.
COMMENT
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPep; Information
on the WormPep database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep CH211-195G24 is
from a CHOR1-211 BAC library
VECTOR: pTRBAC2.1
Clone-derived Zebrafish pUC subclones occasionally display
inconsistency over the length of mononucleotide A/T runs and
conserved TA repeats. Where this is found the longest good quality
representation will be submitted.
Repeat names beginning 'Dr' were identified by the Recon repeat
discovery system (Zhirong Bao and Sean Eddy, submitted), and those
beginning 'dir' were identified by Rick Waterman (Stephen Johnson

```



```

/translate=11
/product="ketoreductase"
/protein_id="CAE17519.1"
/db_xref="GI:40644826"
/translation="MRPGPAGKVALVAGGTRGNGFAVARTIMRGCHVPLANTAHBE
GGAKEAVTALREBGLPGTATLRSDITRPDALPRLLHGTROHHRDLDFVHSAFHP
MNAADAPAPAMATDLTALAPLVYGRPLVRAMAGGPRVYVSTGAQVRPRRYVA
GIKAKALEAOVAVLAVELAPAGITNGSTALDGRARGDPAIMRALAARTPAQLTTT
PDDIAVAVALICSDPARMHGOVITADGGTLTRA"
complement(1585..1836)
/gene="cmmS"
complement(1585..1836)
/gene="cmmS"
complement(1585..1836)
/function="biosynthesis of the polyketide moiety"
/codon_start=1
/translate=11
/product="acyl carrier protein"
/protein_id="CAE17520.1"
/db_xref="GI:40644827"
/translation="MMTAEDLKRIMREGAGODEGVDLGDIDVDPPTDGYDSLALLE
TVSRVERFCGPIDETVGAKTPEGFLHVLNMGSGAS"
2050..3432
/gene="cmmY"
2050..3432
/gene="cmmY"
/function="sugar biosynthesis"
/codon_start=1
/translate=11
/product="NDP-2,3-dehydratase"
/protein_id="CAE17521.1"
/db_xref="GI:40644828"
/translation="MAQATRTADRVAHRYGPRLHSAQAQDRAGGIPGLDEBRRS
GIVKRIPEGEVWKRIIDPTGNLVHDSGRFPSIGLTVRADSGTGHGQPTINDBET
VGRGILVKEIDGVLRFMQAKMKNIGLFOGLPTVOATRSNYTGHHGRPPRYIEYF
NLVMDARIVLDSLOSQDSFLKRNKNVETITDIIVYRDWRMLTIGETHQDLAVP
RTPAIARSDGRLAVLAVOCAGNREYTNWTOPLAQROGIAFIVKRIAGEVOVL
VOGLTOGTPDVAEMAPVQCTVDGHHDLPEKPPYLDVYLAVPDDRVHYSALLSEEG
RFRYRPNRIIVYADETLGDDPPDAVRLMGFSQLKRFQYSNHVVEARTLIACIQLS
WKEFLMANG"
3417..4427
/gene="cmmW"
3417..4427
/gene="cmmW"
/function="sugar biosynthesis"
/codon_start=1
/translate=11
/product="NDP-3-ketoreductase"
/protein_id="CAE17522.1"
/db_xref="GI:40644829"
/translation="MGPLRYGVLCASIAIRKYLPAIASPVLVAVASRTEGKAT
MFAPYGDAVTVGYELLORDVDVAVITPLPALAKATWIALAEGHVGKELTTT
AADARDIVALGRRGLMESFMLHSHVRRVELVAAGRIIGRPSFTAEFTPLP
ADDINRRGLGGGALLDGVYTVRAAOYFCGPGSLVAGAVLMSDPVAGAPPAAGV
DRAGALITSRTTAQLTFGIHAAYOCYRICSEGLSLDRAFTPPNRPVVRLED
KDGTEERALEADHFNANITVAFAGTALGGDLTGHSQILRQADIVERTIRACAPAPR
AA"
4438..5193
/gene="cmmIII"
4438..5193
/gene="cmmIII"
/function="sugar biosynthesis"
/codon_start=1
/translate=11
/product="NDP-4-ketoreductase"
/protein_id="CAE17523.1"
/db_xref="GI:40644830"
/translation="MRITGHTLARSIRPLASHPRVVAFAGVSSGDGVADOPARE
AALITVLGQLDGLRLVYFTSSAGMCPRHAGREDGPVPFSSYGRHLAAEYV
IRNSGADHLIRLAVPGQPHQFLPSLVAOVGRVAVRARGARDIDVRHVSU
LKSLIHEDVROVNINASGEAAVEDIVGLEDRGTIVAEKRTVVPAGPSTERLH
RLVKAELGFCGYRQVIDSYLAVAVSSPA"
6855..8102

```

```

CDS
/gene="cmmOI"
6855..8102
/gene="cmmOI"
/codon_start=1
/translate=11
/product="oxycanase"
/protein_id="CAE17524.1"
/db_xref="GI:40644831"
/translation="MTIDQTEPATGTTDVCVIGGPGAGTLALLLRSGVOYTVR
ATMSREYRGITLQPGIALIDQGLVAPAAARGYWLDSGOLVERERVTLDIAVRL
PAPYRRLISVQPRHIAELDRCPGPGPDYVDOHRTSALVEEDGARGVAGARGP
HVRAGVWVGADGFSFKTRRLAGIGAROPVDFVDFMLPOSSPGRVVRFAAG
RPLVYVPEYPGTVOLGWTLPHKGYAEWATIGETIRSEIARAVAPADRLARVHALS
DPTLDVPRAGARWYVDGLVIGDSHTSHPGAQGINALQDALHPLVLEALRD
GDPSRGLDAYALRRSRDIDQVMKTQTMOKAMLSQGRVAVVRPLAGIYQVTPICP
KITRIHALGNPAARVRSDLFAPG"
8182..8610
/gene="cmmX"
8182..8610
/gene="cmmX"
/function="biosynthesis of the polyketide moiety"
/codon_start=1
/translate=11
/product="cyclase"
/protein_id="CAE17525.1"
/db_xref="GI:40644832"
/translation="MDLALLPOBRIRAGIEGYARHMOULDTRABEWARTTADQTF
ALGRPSPSRGAEALAGASRAISQRADETHRHMHGMLDITVDDGTIVRCYALV
VLPRGDDPLRHACVCEDEVLDVEGGEWKRARVYTRDGA"
complement(8691..9959)
/gene="cmmK"
complement(8691..9959)
/gene="cmmK"
complement(8691..9959)
/function="biosynthesis of the polyketide moiety"
/codon_start=1
/translate=11
/product="ketosynthase"
/protein_id="CAE17526.1"
/db_xref="GI:40644833"
/translation="MRPPLATGMGLVAPNGAQQOAVMAATLRGENGIGRTTRPPAST
PARLAGEIGFDPASAEIPNRLIPOTDEVTRLSTVAAGALADALDPALAEVYMW
TANTAGFPGFQORELEINMSKSGHVSAYQSPAMFTVAQTVSIRGLRGQSVVVA
DGAGGLDALAORRWIRRGCTPLVDSGAVDSLCFSWNTAOITVGRSLTRBDPADAYLP
PBARAGCVLGGGALLVVEDQDSRARARVHQLIDGCTSPDRPGTGRPGLERA
VRALEDAMPADVDVVFADAGSPGPDREEDALAAVFGPHGVPTAPKTMGRLA
AGSGPLDVATALLSPADQVIPPVTVHSEPVQAYRLDVTAPRRVPLRALVLRGHG
NSVAVAVAPPEAGVFPSPGTSARTTIGA"
complement(9956..11224)
/gene="cmmP"
complement(9956..11224)
/gene="cmmP"
/function="biosynthesis of the polyketide moiety"
/codon_start=1
/translate=11
/product="ketosynthase"
/protein_id="CAE17527.1"
/db_xref="GI:40644834"
/translation="MNRVVLTGIVVAPGCTGTAKFWELITSGTATRPISLFDASP
FRSRIABCDPDPALHGLTHKETRRMDRAQFAVAVTDEAVDAGLEFRTDPARGV
ITVSAVCTGTSLEBYTAVSDGRDWMLVDHTYVAPHLVPHLVSSMAAVARVARG
PASVAVGCTSGSLDAVHARGLILEGSADVWAGVADAPASPITLACPAIRAKTTPRG
DDPEHARPPDATRNGFVLGEGSAMVLEVEHARVARGAPVAEITGFPASRCNAHMT
GLRPGDERMEAEIAGALAEARGPDDIDVYNHAGSCTKONDRHETRAFPALGEBHAYA
VPVSSIKSWGSHLGAIEIAKALAAVRHGVPTPALHBPDECDLDVYPLRARH

```

Alignment Scores:

Pred. No.:	1.3e+03	Length:	42074
Score:	116.00	Matches:	68
Percent Similarity:	33.33%	Conservative:	24
Best Local Similarity:	24.64%	Mismatches:	99
Query Match:	7.93%	Indels:	85
DB:	1	Gaps:	11


```

US-09-087-136-1 (1-275) x SGR578458 (1-42074)
Qy 28 SerGlyGlyProLeuProLeuGluSerProSerArglySerLeuThrSerLeuLeuSer--- 46
Db 22898 ACCGATGACGCGCTCCCGCGGAGACCGGTCGACCTGCGAGGAGACGACGAGGCGGA 22839
Qy 47 -----TyrAspProThrValProGluSerProAspMetLeuPheAla 60
Db 22838 TGTCTCTCTCCCGGCGCGCGCGCCGACGCGCTCCCGGACTCCT----- 22794
Qy 61 ArglySerArgLeuGlyAsnLeuLeuThrThrIleIleYshHisProSerGluIleIleGly 80
Db 22793 ---CGGCGCGCGCGCGCGCGAGCTGCACAC-----CGGCTCCG----- 22758
Qy 81 ValLeuProGluAspTyrThrArgAlaAspGluGluProGly-----ArgGlnGlyArg 98
Db 22757 GCACCTCCGACGCGTCAGCGCTCCGGGCGCGGCTCCCGACGACCTCGGCGCGCGCGCGT 22698
Qy 99 ProProGlyArgProArgLysMetProArgHisLeuSerSerThrSerLeuMetGluSer 118
Db 22697 CCACCAAGAAAGACCGCGCGCGCGCGCGAGGATCCCGCGCGAGCTTCACATCC 22638
Qy 119 ProArgLys-ThrMetThrArgAspSerLysIleMetPheGluLeuArgGlyLysProPh 138
Db 22637 CCGGACCGTTGACGCGGTGAGCGGCGGAGCGGCTCCGACGAGGTGGCGCGAGCGCG-- 22580
Qy 138 eGluMetIleAlaGlyArgPheGluGluGluGluTyrSerLeuGlyArgAlaTTPValLysG 158
Db 22579 -----CAQGGCTACCAAGCTCGGCGCGGCTTCGCGCGTGAAGC 22545
Qy 158 YHisMetAsnAsnGluTyrGluProIleLeuAlaGlnArgThrAspTyrAlaProAsnLe 178
Db 22544 ACAC----- 22541
Qy 178 uAlaValAspTyrIleuAlaCysArgGluIleHisArgMetProArgProAspLysSerI 198
Db 22540 -----GATCCCTCATGCGGCGGTGTCGCCATCGCTCCAGCGCGCGCGATCCCGT 22488
Qy 198 ePro-----GluLeuProIleValProSerArgIleAspG 210
Db 22487 AGCTCTGGGCGACGACTCGTCGAACGTGATCTGCGGCTTCGCCGTCGCGAGCGGATCC 22428
Qy 210 uPheAspAlaThrValAspProArgTyrGluThrAspLeuLysAsnGluTyrIleArgH 230
Db 22427 GCGCGGACCAACCGGTGAGCCC-----GCTGAAGAGCTCGTCA 22389
Qy 230 sTrpLysGlnValLysLysGlyTyr-CysAlaHisGlnArgArgThrAlaPro---- 248
Db 22388 C-----ATAGCGGCGCGCTGCGCTGCGCACAGCGGTCCTCCAGCGCGCGGCTAC 22338
Qy 249 -----HisAlaArgSerIleAlaLeuIleAsnL 258
Db 22337 CGGTGCGCTCAGCATCGGCGCGCGCGGTCGACCTCCGCGAGCAGCAGCAGCAGCTTCGATC 22278
Qy 258 yslIleTyrGln-ProGlyGluSerLysThrValGluGlnAla 271
Db 22277 CGCTCCACAGCGCGGCGGTGAGTTTCAACCCCGAGGCGCGG 22236

RESULT 5
BX571714/c 240105 bp DNA linear VRT 01-OCT-2004
LOCUS zebrafish DNA sequence from clone DKEX-16L2 in linkage group 3,
DEFINITION complete sequence.
ACCESSION BX571714
VERSION BX571714.7 GI:53748628
KEYWORDS HTG.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 240105)
AUTHORS Smith,M.

```

```

TITLE Direct Submission
JOURNAL Submitted (01-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zf1sh-help@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Oct 1, 2004 this sequence version replaced gi:50234176.
COMMENT Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zf1sh-help@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
EM, EMBL; SW, SWISSPROT; TR, TREMBL; WP, WORMPEP; INFORMATION
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
zebrafish pUC subclones occasionally display inconsistency over the
length of mononucleotide A/T runs and conserved TA repeats. Where
this is found the longest good quality representation will be
submitted.
Repeat names beginning 'Dr.' were identified by the Recon repeat
discovery system (Zhixiong Bao and Sean Eddy, submitted), and those
beginning 'drr' were identified by Rick Waterman (Stephen Johnson
lab, WashU). For further information see
http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml DKEX-16L2
is from a Zebrafish BAC library
VECTOR: pIndigoBAC-5.
FEATURES
source
1..240105
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEX-16L2"
/clone_1id="DanioKey"
LOCATION/Qualifiers
ORIGIN
Alignment Scores:
Pred. No.: 1.15e+04 Length: 240105
Score: 115.00 Matches: 68
Percent Similarity: 38.46% Conservative: 47
Best Local Similarity: 22.74% Mismatches: 135
Query Match: 7.87% Indels: 49
DB: 5 Gaps: 11

US-09-087-136-1 (1-275) x BX571714 (1-240105)
Qy 10 PheLeuLeuProGluAspGlyAspArgAsnAlaArgGlnAsnAspProLeuIleSerGly 29
Db 147209 TTCTCAATGAAAGAAAGAGAGGTAAAGCATTCGATCATCTGCTTCCTTCA 147150
Qy 30 GlyProLeuProLeuGluSerProSerArgLysLeuThrSerLeuLeuSerTyrAspPro 49
Db 147149 TCTCCTCTCTGTTTCAAGTTCAGATGCTTCATCCGTCGTCCTCGTAAGCACA 147090
Qy 50 ThrValProGluSerProAsp-----MetLeuPheAla 60
Db 147089 CCACACACAGACGACCTTCCTGCTCTTAAGCCGCTCCGATGCTTCCTCCGTTTGGT 147030
Qy 61 ArgLysArgLeuGlyAsnLeuLeuThrThrIleYshHisProSerGluIleIleGly 80

```

Db 147029 TCTGTCCAGTGAAGAGCGCCCTAGCCTCCCTCCCTGCTCCCTTTCAGCCATTCTTACC 146970
Oy 81 ValLeuProGluAAspTyrThrArgAlaAspGluGluProGlyArgGlnGlyArgProPro 100
Db 146969 GTCCCTCCACCTGTGCGGTCTCTTGCGCCCTGATAGGCCCAAAAGATCCAGGTCCTCC 146910
Oy 101 G1YArPProAArgLysMetProAArgH1G1LysSerSerThrSerLeuMetG1LysProAArg 120
Db 146909 CCAATCCCAACAATCCCTCCCAAAATCTCAACCAAAAGTGGCAAAAGCTCTCCCGA 146850
Oy 121 LysThrMetThrArgAspSerLys1LMetPheGluLeuArgGlyLysProPheGlu--- 139
Db 146849 -----ACACCCCTCTGGCGCTGTGTGGCTCAGGTCCAAAATCTCCCTTCACACAT 146799
Oy 140 -----Met1Lea1aGlyArgPheGluGlnGluLysSerLeu-----Gly 152
Db 146798 GCCTTCATGTCTCAAAATGCTTACAGAGATCCGTCCTCCAACTTGCACAACTACAAAGGCG 146739
Oy 153 ArgAlaTTrpValLys-----GlyHisMetAsn--- 161
Db 146738 CGTGACACGATGTGCGCCCTGGACATCAGACATCGGCTCAGCTTCTATCTTCACTCT 146679
Oy 162 -----AsnGluTyrGluProLysArgAlaGlnArgThrAspTyrAlaPro 176
Db 146678 GCTCTGAGAGAAAGAGACGATGAAGAAGAACTGAGACAGATTAAGTAACTCAGGGAA 146619
Oy 177 AsnLeuAlaValAspTyrLeuAlaCysArgGln1LHis1ArgMetProArgProAspLys 196
Db 146618 CAACCTGGAGTGTCCAGCTGTCTG-----CACTGCGCTTGGCCCCCAAGCTTACCTC 146565
Oy 197 Ser1LeProGluLeuProLysValProSerArg1LAspGluPheAspAlaThrValAsp 216
Db 146564 TCTGTT-----CTGGCCGATGTGGCGCTCAGATGACCCG---GAGGCCGATTCAC 146514
Oy 217 ProArgTyrGluThrAspLeuLysAsnGluTyr1LArgHisTrrPylGlnValLysLys 236
Db 146513 TCAGAA---GAACTAGTGAAGCTCTTACGAGGACGAGAAAGCAACAGCTGAGAGAAAGAA 146457
Oy 237 G1YTrpCysAlaHisGlnArgArgArgThrAlaProHisAlaArgSer1LAlaLeu--- 256
Db 146456 GGA-----GACTTTTGTGCCCGGCACCAAGCCCAACCACTCTGTGAT 146415
Oy 256 eAsnLys1LeyTyrGlnProGlyGluSerLysThrValGln1LAlaGlnLysLeu 274
Db 146414 CCAAGGGGTCTATTGTCATGACAGAGACAACTATTCTAAACTCTCGACCTC 146360

RESULT 6
AC146961 181667 bp DNA linear HTG 30-OCT-2003
LOCUS AC146961
DEFINITION Ocolemur garnettii clone CH256-500g8, WORKING DRAFT SEQUENCE, 4
unorderd pieces.
ACCESSION AC146961
VERSION AC146961.1 GI:38044159
KEYWORDS HTG, HTGS PHASE1, HTGS DRAFT.
SOURCE Ocolemur garnettii (small-eared galago)
ORGANISM Ocolemur garnettii
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Strepsirrhini; Galagonidae; Ocolemur.
(bases 1 to 181667)

REFERENCE
AUTHORS Antonellis,A., Ayele,K., Benjamin,B., Blakeley,R.W.,
Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,B.,
Coleman,H., Engle,J., Granite,S., Guan,X., Gupta,J., Haghghgh,P.,
Han,J., Hansen,N., Ho,S.-L., Hu,P., Hurle,B., Idol,J.R.,
Karlin,R., Kwong,P., Latic,P., Larson,S., Lee-Lin,S.-O.,
Legaspi,R., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Mastello,C.,
Maskeri,B., Mcowell,J., Mullikn,J.C., Peguritan,C., Pearson,R.,
Portnoy,M.E., Prasad,A., Reddik-Duque,N., Schandler,K.,
Schuler,M.G., Shah,K., Sison,C., Stantirpop,S., Thomas,J.W.,
Thomas,P.J., Tsipouri,V., Vogt,J.L., Wetherby,K.D., Young,A. and
Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished

REFERENCE 2 (bases 1 to 181667)
AUTHORS Green,E.D.
TITLE Direct Submision
JOURNAL Submitted (30-OCT-2003) NIH Intramural Sequencing Center, 8717
Grosvont Circle, Gaithersburg, MD 20877, USA
COMMENT ----- Genome Center

```

Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@ngri.nih.gov
-----
Project Information
Center project name: fok
Center clone name: 500C08
-----
Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 180010 bases at least Q40
Consensus quality: 180550 bases at least Q30
Consensus quality: 180797 bases at least Q20
Insert size: 185000; agarose-fp
Insert size: 181367; sum-of-contigs
Quality coverage: 12.6ix in Q20 bases; agarose-fp
Quality coverage: 12.86x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of 'N', but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 5956: contig of 5956 bp in length
* * 5957 6056: gap of unknown length
* * 6057 29634: contig of 23578 bp in length
* * 29635 29734: gap of unknown length
* * 29735 80482: contig of 50748 bp in length
* * 80483 80582: gap of unknown length
* * 80583 181667: contig of 101085 bp in length.
*
FEATURES
source
1. 181667
Location/Qualifiers
/organism="Otolemur garnettii"
/mol_type="genomic DNA"
/db_xref="taxon:30611"
/clone="CH256-500G8"
/clone_1fb="CH256"
1. 5956
/note="assembly_fragment"
6057. 129634
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
29735. 80482
/note="assembly_fragment"
clone_end:T7
vector_side:right"
80583. 181667
/note="assembly_fragment"
ORIGIN
Alignment Scores:
Pred. No.: 1.16e+04 Length: 181667
Score: 113.00 Matches: 71
Percent Similarity: 36.71% Conservative: 34
Best Local Similarity: 24.83% Mismatches: 128
Query Match: 7.73% Indels: 53
DB: 2 Gaps: 14
US-09-087-136-1 (1-275) x AC146961 (1-181667)
14 G1uAspG1yAsp-----ArgAna1AArg1nAsnAspProLeuIleSer----- 28

```

Db	155290	GAGGCAAGGAGTAACACCTACCTTAGAAATCGAGACTCCAGACTCTAGTCTTCACTG	155349
Qy	29	gLygIyProLeuProLeuGluSerProSerArgLyLeuThr-----	42
Db	155350	GAAAGGAATCTCACTGGGGATTTCAGACCCAGAGCTCTCAACTGGGCGAGATGAGG	155409
Qy	43	-----SerLeuLeuSerTyThrAspProThrValProGluSerProAspMetLyPheAla	60
Db	155410	AATTTTCCATCCCTAAATTTCTCACTTTTTCCTCCCAAGTACATCAGATGAAGAGAGCC	155469
Qy	61	ArgLyAsArgLeuGlyAsnLeuLeuThrTrtTrtIleLySniHis-~ProSerGluIleIe	79
Db	155470	AAGCACCCAGACAGTGGT-----TCTCGAACAACATCTCCTCCCAAGCTCTTAAAGA	155520
Qy	80	GlyValLeuProGluAspTyThrArgAlaAspGluGluProGlyArgAln--GlyArg	98
Db	155521	GCCTCCAGACCCAGACAG--GCGATAAACCTGAGCAGCCAGGACAGTATGGCCAGA	155577
Qy	99	ProProGlyArgProArgLyMetProArgHis-GluSerSerThrSerLeuMetGlu	118
Db	155578	CCTCCCTCCCGTCCC-----CACCCATCTGGCCCTTCCCAAGCCCTCG	155622
Qy	118	ProArgLyThrMetThrArgAspSerLyIleMetPheGluLeuArgLyLyAspProh	138
Db	155623	TCTCTCCCAAGCGCTAAGCAAG--AGAAAGTTCACAAAGCCCTTTCTTGCCCTG	155679
Qy	138	eGluMetIleAlaGlyArgPheGluGluGluTyTySerLeuGlyArgAlaTrpValLyS	158
Db	155680	GGCTGTACCTTGGGCGCTGCTGCTGACCCAGAAAGTTAGAAATGGGTGCTCAAG	155739
Qy	158	YHisMetAsnAsnGlyTyGluProIleLyAlaGlnArgThrAspTyThrAspAsnLe	178
Db	155740	CAGATCTGTAACTACACACAGAGTGTATTGTCTCAAGGACAGACTAGCCCTTAGCAGAGGT	155799
Qy	178	uAlaValAspTyTyLeuAlaCysArgGlu-----IleHisArgMetProArgProAspLy	196
Db	155800	CCCCACTTCCCTCTTAGATGATGCCAGAGGCGCGCTCTCAACGCGACACCTTACCCTATCC	155859
Qy	196	AserIleProGluLeuAspProIle-----	203
Db	155860	CCCAGGCCCTCCCTTCCCTTCCCAAGGGCGCTCACCATTAAGCCCTGAGAAATCATCA	155919
Qy	204	---ValProSerArgIleAspGluPheAspAlaThrValAspProArgTyGluThrAs	222
Db	155920	ACCAATTCCTAGACAGAGCAGAGCGGCTCTCTCACTCAGCT-----AGAGGAGACAGAA	155973
Qy	222	PLeuLyAsnGlyTyTrtIleArgHisIleTyLySerGlnValLySlyGlyTTPCysAlaHisG	242
Db	155974	CCAAACCCCCACAGAGAGCTGCTACTGCCACCAAGCCCAAGAGCTGTGTGCACACAC	156033
Qy	242	naArgArgArgThrAlaProHisAlaArgSerIleAlaLeuIleAsnLyIleTyGlnPr	262
Db	156034	GAGAGGGCTCAAC---CCAGGTGGCAGA-----TTCACCC	156066
Qy	262	oGlyGluSerLyThr 267	
Db	156067	TGGGAAAGTAAGTACGCC 156082	
RESULT 7			
AE004648		11031 bp	DNA linear BCT 19-FEB-2003
LOCUS			
DEFINITION			Genome. Pseudomonas aeruginosa PAO1, section 209 of the complete
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
			1 (bases 1 to 11031)
			Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
			Hickey,M.T., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,

TITLE	JOURNAL MEDLINE PUBLISHED REFERENCE AUTHORS
agrou, M., Garber, R.L., Goltzy, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A.A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H., Wong, G.K.-S., Wu, Z., and Paulsen, I.T.	Complete genome sequence of <i>Pseudomonas aeruginosa</i> PAO1, an opportunistic pathogen Nature 406 (6799), 959-964 (2000)
20437337	
10984043	
2 (bases 1 to 11031)	
Stever, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warren, P., Hickey, M.J., Brinkman, F.S.L., Heflinger, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltzy, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A.A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H., Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reiter, J., Sater, M.H., Hancock, R.E.W., Lory, S. and Olson, M.V.	Direct Submission Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA 3 (bases 1 to 11031)
REFERENCE	
AUTHORS	
CONSRM	
TITLE	
JOURNAL	
COMMENT	

This represents the February 3, 2003 version of the continually updated, reviewed, *Pseudomonas aeruginosa* PAO1 genome annotation, from PseudoCAP (see <http://www.pseudomonas.com> for latest updates and links to alternate annotations). PseudoCAP is coordinated by Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert E.W. Hancock (University of British Columbia, Canada). We welcome submission through www.pseudomonas.com of any proposed changes.

'Protein name confidence' is used to rate our confidence of the accuracy of the protein name.
Class 1: Function experimentally demonstrated in *P. aeruginosa*.
Class 2: Function of highly similar gene experimentally demonstrated in another organism (and gene context consistent in terms of pathways its involved in, if known).
Class 3: Function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene.
Class 4: Homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences.

```

location/qualifiers
1. .11031
   /organism="Pseudomonas aeruginosa PAO1"
   /mol_type="genomic DNA"
   /strain="PAO1"
   /db_xref="taxon:208964"
124. .1446
   /locus_tag="PA2214"
124. .1446
   /locus_tag="PA2214"
124. .1446
   /locus_tag="PA2214"
   /note="Protein name confidence: Class 3 (function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene)"
   /codon_start=1
   /transl_table=11
   /product="Probable MFS transporter"
   /protein_id="AA05602.1"
   /db_xref="GI:9948238"
   /translation="MSAMNDHARLSSAITSRLPLFLVLYWLAIFRANVGVARAFIADSLGDAALFAFGASITFTGYCEVPSNLMHRIGARWLICRIWTGVSAAMPRADDMTPTVYLRFLIGVAAGCFPIIIILYTYWPPQRRAQALDFYGCPLAVLTGGPISLLSEPHGIPGLANNQMPLFVEHGVLASIVGAIAFYLVDRPDACMGDEOKRALBELAEEDTRKQARGPHQLFRSGVGLAKCLVVFITYOMSVGVGFPLPRISAE

```

LDGQVGLNGLITAIPIWVCALVTRVLTRODRSGGROHLAVMLGMAAGIASAL
GNLGLAVIAPCFRAGSFVSQVLPFWTLPTGVLGAAASGIALINSIGTGFVAPFL
KTMESQPAFPRAAGMFAALVGLIGACILARLKTSSGSSVPLARQNG"

gene
CDS

/locus_tag="PA2215"
1477..2652
/locus_tag="PA2215"
/note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAG05603.1"
/db_xref="GI:9948239"

/translation="MKIKSVTRVFEKGVNPPQAHFCMASDIIPEKDGMSFRF
HGMIVAEITDDGLVGINCALAPRAKEIVLDYLAICIGEDPFNEEYIMQMRRT
HANGKIGMAISAVDLAIWIDMKAVNKPVPKLGRTKEKITWYASKIYANDNL
AFLEAQGYLNGPTALKMFYGYDGPDKMGKRNIEQVRLRELAPDIDIMLEGY
GMTLEVARMLPLTAEFEPRMLEBPVADIIEGYVELKMGIMPTSGSHEFTGFK
DLERPAVDVIOYDTRVGGITARKINMAEASVPVIPHAGOLHNHYLTMASTAP
MAEFPVDEVENELFYVFKGEPQPVGCIYQLDHPKGLGLEISEHDKDFITTE"

gene
CDS

2680..3675
/locus_tag="PA2216"
2680..3675
/locus_tag="PA2216"
/note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAG05604.1"
/db_xref="GI:9948240"

gene

/translation="MKLIQFEDRAGORRYVGEAGIOTVARGSTRRELGLAIRAGS
GHADYLRRSGRPGDYGLEERGVLPDLDDPAHCVSTGITHIGSATDRMH
OONOGDETALPTMRLFRWLEGGKRPAGQVAGQEMFKDGQGIIVAGADFLPAP
ABDAGEBPETVLVYIIGDRRPRYLGVALGNEFSDLMERRYLYLAHASKIRACCYGP
ELRVGLPRLHIGESRIILRDGRVILQQAFLSGEDMCHSLNLEHYHFKVQFLRPGD
VNAVHFGATLSPFADGIRKAPGDVPEIAMEERGAFLRNGIAAVETALPRGRVAL"

CDS

3804..5387
/locus_tag="PA2217"
3804..5387
/locus_tag="PA2217"
/note="Protein name confidence: Class 3 (function proposed based on presence of conserved amino acid motif, structural feature or limited sequence similarity to an experimentally studied gene)"
/codon_start=1
/transl_table=1
/product="probable aldehyde dehydrogenase"
/protein_id="AAG05605.1"
/db_xref="GI:9948241"

/translation="WTALIGHNPIGARSAGTLEFLOSLDAAGEALPYRVOATPE
VDAAEAASAVPHYROLPASRAPELDTIAEALDDDDPAVCEITAPTRTIOG
ERSRSTGQLRFAVLIRGRDGHGARDIRAREKPLPRVDLRQRTIGLGVAFGAG
FLPASTGQPTAAALAGCPVFRASGHATAERAAVAILRAEMTSINPAGVPMI
YGGVGERLVHPAIQAVFTGSLKGRALCDMAARAQGPVPAEMTSINPVLVLA
ALKRGEAVDELASVVLGGQFCNPGVITGIRSAQSFALERFARAMDPOATM
LNTGTLSEYERGLAALAHAPRVRLHAGOPGROARPOLROADVSLLEGEDELQEV
FGPASVVEVADHAELKRALHGLHGLTATLIAEAPDLAPDLVPLEBKARLLIN
GPTGVECDAMVHGPPATSDARGISVGTIALIDRLRVCQTQNTPDALPLALDQ
NPLGIARLVGIYRAVA"

gene

CDS

complement(5526..6629)
/locus_tag="PA2218"
/note="synonym: yode"
complement(5526..6629)
/locus_tag="PA2218"
/note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)"
/codon_start=1
/transl_table=1

/product="hypothetical protein"
/protein_id="AAG05606.1"
/db_xref="GI:9948242"

gene

CDS

/translation="METKSNRARSRGALRGAVLAQALMALVCCQTSPPAATSSNTG
GMMOLOTQGWMDKTFPLSAKVERKTFPARKVITTLAADLYPRKNGDRLPALVIG
GPGVAKEQSGLTAQWAEKRGFTYLADEBSYGESGQPPNVAASPDIINDEFAAVD
FISLPEVNRERIEVIGICMGVMAALNAVAKVAVVISTYMDTRVMSKGVNDV
TLEQRTLEQLGQQRWMDAESCTPAYQPPYNEIKGEAQFLVDHYDWTMPGYHPR
AVNSGMWMTTTPLSFMMRPILTYIKSISPRPILHIGERAHRSYFSETVYAAAEK
ELLIVPGASHVDLYDRLDRIPEFDRIAPFPEHL"

complement(6723..7931)
/gene="opdB"
/locus_tag="PA2219"
complement(6723..7931)
/gene="opdB"
/locus_tag="PA2219"

gene

CDS

/note="Protein name confidence: Class 1 (function experimentally demonstrated in P. aeruginosa); Subcellular localization: inner membrane protein; Subcellular localization confidence: Class 1"
/codon_start=1
/transl_table=1
/product="membrane protein OpdB"
/protein_id="AAG05607.1"
/db_xref="GI:9948243"

Alignment Scores:
Pred. No.: 530
Score: 112.00
Percent Similarity: 31.56%
Best Local Similarity: 24.33%
Query Match: 7.66%
DB: 1
Gaps: 14

US-09-087-136-1 (1-275) x AE004648 (1-11031)

QY	13	ProgluaspGlyAsparghsma1aargglnaenaspProleuileSerGlyGlyProleu	32
DB	810	CCGCAAAACGACGACGCGCCGACAGG	836
QY	33	ProleuGluserProSerarGlyserThrSerleuSerTyraSpProThrValPro	52
DB	837	-----	842
QY	53	GluserProaspMetCysPhe1aarglysarGleuglyAenleuSerThr11lels	72
DB	843	CAGCCGCT-----GCGCAGCGCCAGAGTGCCTGAAGTTCTCCTGCTCTACTT	890
QY	73	H1SH1SPProSerGlu11le-----GlyValleuProgluaspTyThrArgAlaasp	90
DB	891	CACCATCCAGATGACCGTACGCGGTGTCTACTCGCCAGC-----	935
QY	91	GlugluPProGlyArg---GlnglyArgProProgluArgProarglySerMetProArgHis	109
DB	936	-----CCGATCCGTAGCTCTCTGACGCGCCAGATGCGCTGACGCGCTGATCAC	989
QY	110	GluserSerThrSerleuMetGluserProArglySerThrMetThrArgAspSerlyle	129

Db	990	CGCGATTCCCTGGGTCGTGGCCCTGATGTAAACGGCCCT	-----	1028					
Qy	130	MetPheGluLeuArgGlyAlaTTPValIysGlyHisMetAsnAsnGluTyrGluProIleLeuVal	149						
Db	1029	-----GGTGAACCCG-----	-----CCAGCGCCGACCGGACGGCCAGCA	106					
Qy	150	SerLeuGlyArgAlaTTPValIysGlyHisMetAsnAsnGluTyrGluProIleLeuVal	169						
Db	1062	TCGCAGCTCGCGTGGCCAGTCTGAGCATGGCCGACGCGGATATCGCCGCTCGAGCGCT	1121						
Qy	170	GlnArgThrAspTyrAlaProAsnLeuAlaValAlaIysTyrLeuAlaGlySerGluIle	188						
Db	1122	GAGCCG-----CAACCTGGCGCTGGCGGTATCGCTTCTGCTTGGCCGCGCTCGG	1177						
Qy	189	-----HhArgMetPro-----	192						
Db	1173	CTTGCTCTCGGTCAGACCGCGTGTCTTGGACCTTCTCCACCGGCTATGAGCGTGGCG	123						
Qy	193	-----ArgProAspIysSerIleProGluLeuProIleValProSerArg	207						
Db	1233	CGCGCCACGCGGCACTGGCCCTGAT-----CAACTCTCGGCGCAACTCGGCGG	1280						
Qy	208	IleAspGluPheAspAlaThrValAspProArgTyrGluThrAspLeuIysAsnGluTyr	227						
Db	1281	CTTGCTCGCGCGCAACTCAA-----GACCT-----	130						
Qy	228	IleArgHisTTPValIysGlyHisMetAsnAsnGluTyrGluProIleLeuVal	244						
Db	1308	-----CATGGAAACCGATTCGCCGACCGCGCGCATGTTCGCCCTGGCCGCGCT	1361						
Qy	245	ArgThrAla	247						
Db	1362	CGGCTGCT	1370						
RESULT 8	BT005099	1045 bp	mRNA	linear	PLN 04-MAR-2003				
LOCUS	BT005099	Arabidopsis thaliana clone U50283 unknown protein (Atg232840)	mRNA						
DEFINITION	complete cds.								
ACCESSION	BT005099	GI:28827575							
VERSION	BT005099.1								
KEYWORDS	FLI CDNA.								
SOURCE	Arabidopsis thaliana (thale cress)								
ORGANISM	Arabidopsis thaliana								
REFERENCE	Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M.,								
AUTHORS	Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C.,								
	Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y.,								
	Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M.,								
	Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shim, P.,								
	Southwick, A., Tripp, M.G., Wu, T., Shinzaki, K., Davis, R.W.,								
	Ecker, J.R. and Theologis, A.								
	Arabidopsis Open Reading Frame (ORF) Clones								
	Unpublished								
	2 (bases 1 to 1045)								
	Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M.,								
	Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C.,								
	Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y.,								
	Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M.,								
	Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shim, P.,								
	Southwick, A., Tripp, M.G., Wu, T., Shinzaki, K., Davis, R.W.,								
	Ecker, J.R. and Theologis, A.								
	Submitted (04-MAR-2003) Plant Gene Expression Center, 800 Buchanan								
	Street, Albany, CA 94710, USA								
	The RIKEN Genomic Sciences Center (GSC) members carried out the								
	collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN								
	Arabidopsis Full-Length cDNA') : Seki, M., Narusaka, M., Ishida, J.,								
	Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,								
TITLE									
JOURNAL									
REFERENCE									
AUTHORS									
COMMENT									

The Sakl, Stanford, PGEC (SSP) Consortium members constructed and sequenced the PUN1 (ORF) clones using the RAPL cDNAs: Yamada,K., Chan,M.M., Chang,C.H., Dale,J.W., Hsuan,V.W., Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.C., Tortumt,M., Wong,C., Yu,G.G., Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.J., Nguyen,M., Palm,C.J., Shinn,P., Southwick,A., Tripp,M.G., Wu,T., Davis,R.W., Ecker,J.R. and Theologis,A.

Yamada,K. (SSP/PSEC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki K. (RIKEN GSC) and Theologis.A. (SSP /PGEC) contributed equally to this work as PIs.

Annotation based on July 2002 version of the Arabidopsis genome submitted to Genbank.
Location/Qualifiers

1..1045
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="2"
/clone="U50283"
/ecotype="Columbia"
/note="This clone is in pUNI 51."
1..1045
/gene="At2g32840"
1..1014
/gene="At2g32840"
/codon_start=1
/evidence=experimental
/product="unknown protein"
/protein_id="AAO50632.1"
/db_xref="GI:28827576"
translation="MSTPDNPNPNPNPSVAVSNTPIYTASESPPTOPMTVTTPSSPOPOPASSVAIALPHHNPONTYNPLIRSNYSNSHPQPPHDSSIILPESSGRFPTRPRQNNSVADPVGSPSGGYTREGPYCYGHGFNLDPMPMRHPNQSQPOLDSGMKGAPHLQPRAPTSPSTLSDSHKAKSRDALVLVKRKHITGGASLYSLCRSMWRNGAHBEIKPORIDMMTCLEPLVDKTETSLPDLVEALCIENDEESVKHLSESDLTKRHIDRAKVARIEREELKIARYKARLATLLPFGEQRNE"

3'UTR 1015..1045
/gene="At2g32840"

ORIGIN

Alignment Scores:

Pred. No.:	37.4	length:	1045
Score:	111.50	Matches:	62
Percent Similarity:	38.79%	Conservative:	40
Best Local Similarity:	23.57%	Mismatches:	90
Query Match:	7.63%	Indels:	71
DB:	8	Gaps:	15

US-09-087-136-1 (1-275) x BT005099 (1-1045)

Dy 18 ArgAenAlaArgGLnAsnaBProLeuileSerGlyLyProleubProleugluSerPro 37
||| ||||| |
331 CGACCGGTAGACAGATCCATTCCGCCGCATG-----COAGTTGGTAAGTCTT 381
Dy 38 SerAglyLSLeuthrSerleuleusertyYaapPrrothValPrGoluberProabpMet 57
||| ||||| |
382 AGTCCAGG-----GGGTACTCTCCCGAGGCCCA-----4111
Dy 58 LysPhelialarglysleughlsanleuleuthThrlIeluyhlshlProsersglu 77
412 -----GTTTAGCATCACCCATCGGCCCA 435
Dy 78 IlellelgylValleupro-----GlubeptyrThrArgAlaApGelugluProglyArg 95
::: ||||| :
Db 436 TTCGTGGAATTTTGATCATGATCATGATCATGAGAAGCTGCACATCCTCAGATCAA 495
GlnGlylaArgProGly-----ArgProArg 104
||| :::||||| :

Db 496 CAATCTCCTCAACTGGTTCGCGTCAATAGAAAGAGTTCCTCATTTTTCGACCTCGG 555
Qy 105 LysMetProArgHisGluSerSerThrSerLeuMetGluSerPro-----ArgLysThr 122
Db 556 GCTACTCTCT-----TCTCCAACTTCATTAATTTAGAACAGAGTGGCATTAAGAACGCC 606
Qy 123 MetThrArgAspSerLysIleMetPheGluLeuArgGlyLysProPheGluMetIleAla 142
Db 607 AGGAGCCGGAGATGATGCTGTGCTCTC-----GTTAGAAAAAGAGGTTAGTAACCGAG 663
Qy 143 GlyArgPheGluGluGluLysSerLeuGlyArgAlaTyrValIleGlyHisMetAsnAsn 162
Db 664 GGA-----GCTTCTTAATCTGCTGTGTGCATCTTGTTAGA-----AATGCT 708
Qy 163 GluTyrGluProIleLysAlaGlnArgThrAspTyrAlaProAsnLeuAlaValAspTyr 182
Db 709 GCTCATGAGGGAATTAAGCCGACGCA-----ATTGATATG 744
Qy 183 LeuAlaCysArgGluIleHisArgMetProArgPro-----AspLysSerIlePro 199
Db 745 ATGACATGT-----TTGCCAAAGCCTTTAATCTGTGACAAAGCGAGACA 789
Qy 200 GluLeuPro-----IleValProSerArgIleAspGluPheAsp-----Ala 213
Db 790 AGTTGCCAAAAAGATTAGTTAGTGAAGACAAATTTGCGAAGAACAAAGAGAGCAGGAA 849
Qy 214 ThrValAspProArgTyrGluThrAspLeuLysAsnGluTyrIleArgHisTyrLysGln 233
Db 850 TCTGTGAACAATTTGTCAATCTGATCTTTGAAAGACATATGACCGAGCTAAGAG 909
Qy 234 ValLysLysGlyTyrCysAlaHisGlnArgArgThrAlaProHisAlaArgSerIle 253
Db 910 GTCCGCGCTCAGATTAGAGAGACGCTTAAAGAAATCGCAGGTACAGGCAAGATTG 969
Qy 254 AlaLeuIle 256
Db 970 GCTCTTCTT 978

RESULT 9
AK117550
LOCUS AK117550 1324 bp mRNA linear PLN 14-FEB-2004
DEFINITION Arabidopsis thaliana At2g32840 mRNA for unknown protein, complete
cds, clone: RAFL17-20-B19.
ACCESSION AK117550
VERSION AK117550.1 GI:26450184
KEYWORDS FLI cDNA; CAP trapper.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatopsida; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosid II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Seki,M., Iida,K., Satou,M., Sakurai,T., Akiyama,K., Ishida,J.,
Nakaijima,M., Enju,A., Kamiya,A., Narusaka,M., Carninci,P.,
Kawai,J., Hayashizaki,Y. and Shinozaki,K.
TITLE Arabidopsis thaliana full-length cDNA
JOURNAL Published Only in Database (2002)
AUTHORS 2 (bases 1 to 1324)
TITLE Direct Submission
JOURNAL Submitted (25-NOV-2002) Motoaki Seki, RIKEN Genomic Sciences
Center; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa
230-0045, Japan (E-mail:mseki@gs.riken.go.jp,
URL:http://pfweb.gsc.riken.go.jp, Tel:81-45-503-9625,
Fax:81-45-503-9586)
COMMENT An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al. (1998) Plant J. 15:707-720).
Seki et al. (2002) Science 296:141-145). cDNA cleaved with BamHI
and XhoI was ligated to modified lambda PhiC-1-E vector (Carninci et
al. (2001) Genomics 77:79-90) digested with BamHI and SalI.
This clone is in a modified pBluescript vector.

FEATURES
source
Location/Qualifiers
1..1324
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="2"
/clone="RAFL17-20-B19"
/ecotype="Columbia"
/note="common name: thale cress"
1..1324
/gene="At2g32840"
14..1127
/gene="At2g32840"
/codon_start=1
/product="unknown protein"
/protein_id="PAC42211.1"
/db_xref="GI:26450185"
/translation="MSTPDPNPNPNPVSVAVSTPIYVASEPSVYQPNVTIPSSQ
FOPTPASSTRALAPVLRHHPHONTNPLPIKRSNVTNSPHOPHPDPSLLTPFG
SSGRGFPTRPVRONSNVADPVGSPSGGTPRGVYHHGQFVSLDPMNQPMRA
HPQVQSPDQSGSHMKGVPHLPAPRPSPTSLDNGSHKAKSRDPLVLRKRYR
ITEGASLVSICRSWLRNGAEGIKPORIDMMTCLPKPLPVKETSIPKLVSEATCE
EDKDEESVNHLSDDLKRHHIDRAKVRARIRERIKRIARVYARALALLPPREGCC
RNE"
ORIGIN
Alignment Scores:
Pred. No.: 49.2 Length: 1324
Score: 111.50 Matches: 62
Percent Similarity: 38.78% Conservative: 40
Best Local Similarity: 23.57% Mismatches: 90
Query Match: 7.63% Indels: 71
Gaps: 15
US-09-087-136-1 (1-275) x AK117550 (1-1324)
Qy 18 ArgAsnAlaArgGlnAsnAspProLeuIleSerGlyLysProLeuGluSerPro 37
Db 444 CGACCCGTTAGACGAATTCCAATTCGCGCTGAT-----CCAGTGTAGTCTCT 494
Qy 38 SerArgLysLeuThrSerLeuLeuSerTyrAspProThrValProGluSerProAspMet 57
Db 495 AGTCCAGCT-----GGGTATATCTCTCGAGGCCA----- 524
Qy 58 LysPheAlaArgLysArgLeuGlyAsnLeuLeuThrThrIleLysHisProSerGlu 77
Db 525 -----GTTTACGGATACCAACCATGGCCAA 546
Qy 78 IleIleGlyValLeuPro-----GluAspTyrThrArgAlaAspGluGluProGlyArg 95
Db 549 TTGCTGCGAATTTGATTCAGTAATCAATCAGTTCATGAGAGTGCATCTCAGAAATCAA 608
Qy 96 GlnGlyLysProProGly-----ArgProArg 104
Db 609 CAATCTCCTCAACTGGTTCGCGTCAATGAAAGAGTTCCTCATTTTTCGACCTCGG 668
Qy 105 LysMetProArgHisGluSerSerThrSerLeuMetGluSerPro-----ArgLysThr 122
Db 669 GCTACTCTCT-----TCTCCAACTTCATTAATTTAGAACAGAGTGGCATTAAGAACGCC 719
Qy 123 MetThrArgAspSerLysIleMetPheGluLeuArgGlyLysProPheGluMetIleAla 142
Db 720 AGGAGCCGGAGATGATGCTGTGCTCTC-----GTTAGAAAAAGAGGTTAGTAACCGAG 776
Qy 143 GlyArgPheGluGluGluLysSerLeuGlyArgAlaTyrValIleGlyHisMetAsnAsn 162
Db 777 GGA-----GCTTCTTAATCTGCTGTGTGCATCTTGTTAGA-----AATGCT 821
Qy 163 GluTyrGluProIleLysAlaGlnArgThrAspTyrAlaProAsnLeuAlaValAspTyr 182
Db 822 GCTCATGAGGGAATTAAGCCGACGCA-----ATTGATATG 857

Qy 183 LeuAlaCyArGluIleHieArGMeProArGPro-----AspLysSerIlePro 199
Db 858 ATGACATGT-----TTGCCAAAGCCTTTACTCTGTGACAGACGAGACA 902
Qy 200 GluLeuPro-----IleValProSerArGileAspGluPheAsp-----Ala 213
Db 903 AGTTTGCCTCAAAAGATTGTTGAGAGACATTTCCGAGAAACCAAGAGACGAGAA 962
Qy 214 ThrValAspProArGlyTrGluThrAspLeuLysAsnGluTrpIleArgHisTrpLysGln 233
Db 963 TCTGTGAACACATTTGTCTGCAATCTGATCTTTTGAAGAACATATGACCGGACTAAGAG 1022
Qy 234 ValLysLysGlyTrpCyAlaHisGlnArgArGTrGThrAlaProHisAlaArgSerIle 253
Db 1023 GTCCGCGCTGCATTTGAGAGAAACGGTTAAAGAAATGCGAGGTACAGGCAAGATTG 1082
Qy 254 AlaLeuIle 256
Db 1083 GCTCTTCTT 1091
RESULT 10
AL645845/c 109446 bp DNA linear ROD 24-OCT-2002
LOCUS Mouse DNA sequence from clone RP23-338U18 on chromosome 11,
DEFINITION complete sequence.
ACCESSION AL645845
VERSION AL645845.14 GI:24394922
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 109446)
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (23-OCT-2002) Wellcome Trust Sanger Institute, Hinxton,
JOURNAL Cambridgehire, CH10 1SA, UK. E-mail enquiries: clonerequests@sanger.ac.uk
On Oct 25, 2002 this sequence version replaced gi:23337187.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; SW,
SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
from the RPCR-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
Location/Qualifiers
1..109446
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"

/chromosome="11"
/clone="RP23-338U18"
/clone_1fb="RPCR-23"
ORIGIN
Alignment Scores:
Pred. No.: 8.95e+03 Length: 109446
Score: 111.00 Matches: 75
Percent Similarity: 33.33% Conservative: 29
Best Local Similarity: 24.04% Mismatches: 103
Query Match: 7.59% Indels: 105
DB: Gaps: 17
US-09-087-136-1 (1-275) x AL645845 (1-109446)
Qy 10 PheLeuLeuProGlu---AspGly-----AspArgAsnAlaArgGlnAspPro 25
Db 32256 TTCTGTGAAACAGAGGGATGGGTGAGAGACATCACAGGCTTCTCAACCCGACCG 32197
Qy 26 LeuIle---SerGlyGlyProLeuProLeuGluSerProSerArgLysLeuThrSerLeu 44
Db 32196 TTACTGGGCTGGGCGCGGCACTTCCGCGCTCTCCCTCCAGGCCAGCG----- 32149
Qy 45 LeuSerTrpAspProThrValProGluSerProAspMetLysPheAlaArgLysArgLeu 64
Db 32148 -----CCCTCGGTCTCTCCCGCGCGCTCAGGCCAGCGTCCAGCGGCGCTG 32101
Qy 65 GlyAsnLeuLeuThrThrIleLysHisPheSerGluIleGlyValLeuProGlu 84
Db 32100 GGGAGCATG-----TCTCTCCACCCCCAC-----CCACCC 32071
Qy 85 AspTrpThrArgAlaAspGluGluProGlyArgGlnGly----- 97
Db 32070 TAGAAACCGACAGAGGGGCGCCCGCGTAAATCCCGCGGTTCCAGCGGACCC 32011
Qy 98 -----ArgProProGly----- 101
Db 32010 TCGGTGCCCGAAACCCGCTGACGACACAGAGGCGGTATCCCGCGCGCATTAATGC 31951
Qy 102 -----ArgProArgLysMetProArgHisGluSerSerThrSerLeuMetGlu 117
Db 31950 AGCAGTTGTTTCAAGCGCGCGCGGATCTTCCCGACCTCCCGCGCATCTCGTCTC 31891
Qy 118 SerProArgLysThrMetThrArgAspSerLysIleMetPheGluLeuArgGlyLys--Pr 137
Db 31890 CCTCCAGGCGAGACTCTTCAAGACGCGTC-----CTAGAGCCAGCGCGGTAAAC 31840
Qy 137 oPheGluMetIleAlaGlyArgPheGluGluGluTrpSerLeuGlyArgAlaTrpVal-- 156
Db 31839 ATTCCGAGCCACCGCGCTCGCGCCAGAG-----CTCTCGGTTCTGGGTAAAC 31789
Qy 157 -----LysGly-----HisMet 160
Db 31788 CCTCGTCACTGCGCGCTGCGCCCGAGGCGCTTTTATTTTAAACCACTGTCACGT 31729
Qy 160 t-----AenAsnGluTrpGluProIleLysAlaGlnArgThrAs 173
Db 31728 CCGGCTCTTTGGAGAGAGATCGGTCACCGTTTCCGAGGAGGAGCGAAGAGCTCCG 31669
Qy 173 pTyAlaProAsnLeuAlaValAspTrpLeuAlaCyArGluIleHisArgMetProAr 193
Db 31668 GTTCGCTCT-----CTTCGCGCAACCCCGCAG 31642
Qy 193 gProAspLysSerIleProGluLeuProIleValProSerArgLysLeuPhe----- 211
Db 31641 GCTCGGC-----CCGATGTCGTCGTCGAGGTTCCGCGGTTTCGCCCC 31600
Qy 212 -----AspAlaThrValAspProArGlyTrpGluThrAspLeuLys-----AenG 226
Db 31599 TGCAGAGTCGCGCGCTTACGGGGAACCGAGAGAGACGATCGTGAAGACAGAGAG 31540
Qy 226 uTyTrIleArgHisTrpLysGlnValLysGlyTrpCyAlaHisGlnArgArgTrpTh 246

Db 31539 GAAGCGAGAAATGGCGTCCACGGGTAAGCGAGTGGATCGCGGTCCCGCGGCGACA 31480
Qy 246 rAlaProHisAlaArgSerIleAlaLeuIleAsn 257
Db 31479 GCTGGGACACCAGACTGGAGGCGCTGCTCT 31446

RESULT 11
AC005528
LOCUS
DEFINITION Mus musculus clone rp21-493n6 map 11a2 strain 129S6/SvEvTac,
complete sequence.
AC005528
VERSION
KEYWORDS AC005528.43 GI:42491425
SOURCE HTG.
ORGANISM Mus musculus (house mouse)

REFERENCE
AUTHORS Eukariyoti; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL 1 (bases 1 to 199326)
Fang, F., Ying, F., Pan, H., Dumanski, J. and Roe, B. A.
REFERENCE
AUTHORS Mus musculus PAC clone rp21-493n6 in MDR Region
TITLE Unpublished
JOURNAL 2 (bases 1 to 199326)
Fang, F., Ying, F., Pan, H., Dumanski, J. and Roe, B. A.
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (17-APR-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
JOURNAL 3 (bases 1 to 199326)
Fang, F., Ying, F., Pan, H., Dumanski, J. and Roe, B. A.
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (23-MAY-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
JOURNAL 4 (bases 1 to 199326)
Fang, F., Ying, F., Pan, H., Dumanski, J. and Roe, B. A.
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (07-FEB-2004) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
JOURNAL 5 (bases 1 to 199326)
Fang, F., Ying, F., Pan, H., Dumanski, J. and Roe, B. A.
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (10-FEB-2004) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
JOURNAL 6 (bases 1 to 199326)
Fang, F., Ying, F., Pan, H., Dumanski, J. and Roe, B. A.
COMMENT On Feb 10, 2004 this sequence version replaced gi:42475600.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNDR

FEATURES
Source
Location/Qualifiers
1. 199326
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="129S6/SvEvTac"
/db_xref="taxon:10090"
/map="11a2"
/clone="rp21-493n6"
/id="RP21-493n6" - 21 Female (129S6/SvEvTac) Mouse PAC
Library"

ORIGIN
Alignment Scores:
Pred. No.: 1.79e+04 Length: 199326
Score: 111.00 Matches: 75

Percent Similarity: 33.33%
Best Local Similarity: 24.04%
Query Match: 7.59%
DB: 10
Gaps: 17

US-09-087-136-1 (1-275) x AC005528 (1-199326)

Qy 10 PheLeuLeuProGlu---AspGly-----AspArgAsnAlaArgGlnAsnAspPro 25
Db 92810 TTTCTGTTGAACAAGGAGATGGGTTGAAGACGATCAAGAGCTTTCTCAACCCGACCG 92869

Qy 26 LeuIle---SerGlyGlyProLeuProLeuGluSerProSerArgGlyLeuThrSerLeu 44
Db 92870 TTACTGGGATGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 92917

Qy 45 LeuSerGlyAspProThrValProGluSerProAspMetGlySerAlaGlyValGluLeu 64
Db 92918 -----CCCTGGTCCCTCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 92965

Qy 65 GlyAsnLeuLeuThrThrIleValHisProSerGluIleIleGlyValLeuProGlu 84
Db 92966 GGAGAGATG-----TCTCCACACCCAC-----CAACC 92995

Qy 85 AspGlyThrArgAlaSerGluGluProGlyArgGlnGly----- 97
Db 92996 TAGAAACCGACGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 93055

Qy 98 -----ArgProProGly----- 101
Db 93056 TCGGTGCGCGAACCCTGACGACGACGACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 93115

Qy 102 -----ArgProArgGlyMetProArgHisGluSerSerThrSerLeuMetGlu 117
Db 93116 AGCAGTTGTTTCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 93175

Qy 118 SerProArgGlySerMetThrArgAspSerGlyIleMetPheGlnLeuArgGlyLys-Pr 137
Db 93176 CTTCCAGGCGAGACTTTTACAGACGCTCC-----CTAGAGCGACCGGCGGCGGCGGCGG 93226

Qy 137 PheGluMetIleAlaGlyArgPheGluGluGluGluGluGluGluGluGluGluGluGluGlu 156
Db 93227 ATTCCAGCCACCGCGGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 93277

Qy 157 -----LysGly-----HisGly----- 160
Db 93278 CTTGCTCACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 93337

Qy 160 t-----AsnAsnGluGluGluProIleValAsnAlaArgThrAs 173
Db 93338 CCGGCTCTTGGGAGAGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 93397

Qy 173 PThrAlaProAsnLeuAlaValAspTyrLeuAlaCysArgGluIleHisArgMetProArg 193
Db 93398 GTTCGCTCTCT-----CTTCGCGCACCCCGAG 93424

Qy 193 GProAspLysSerIleProGluLeuProIleValProSerArgIleAspGluPhe----- 211
Db 93425 GCCTGGC-----CCATGTATACCTTCGAGGTGGCGGCTTTCGCGCC 93466

Qy 212 -----AspAlaThrValAspProArgGlyGluThrAspLeuLys-----AsnG1 226
Db 93467 TCGCAGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 93526

Qy 226 uTyrIleArgHisTTrpLysGlnValLysGlyTTrpCysAlaHisSerIleArgGlyGlyGly 246
Db 93527 GAAGCGAGAAATGGCGCTCCACGGGTAAGCGAGTGGAGCGGCTCCCGCGGCGACA 93586

Qy 246 rAlaProHisAlaArgSerIleAlaLeuIleAsn 257
Db 93587 GCTGGGACACCAGACTGGAGGCGCTGCTCT 93620

RESULT 12
SLIINC/c

LOCUS SLIINC 36270 bp DNA linear BCT 15-MAY-1996
DEFINITION S.lincolnensis (78-11) lincomycin production genes.
ACCESSION X79146
VERSION X79146.1 GI:499194
KEYWORDS gene cluster; lincomycin biosynthesis; lincomycin resistance; lmbA gene; lmbB gene; lmbB2 gene; lmbC gene; lmbD gene; lmbE gene; lmbF gene; lmbG gene; lmbH gene; lmbI gene; lmbJ gene; lmbK gene; lmbL gene; lmbM gene; lmbN gene; lmbO gene; lmbP gene; lmbQ gene; lmbR gene; lmbS gene; lmbT gene; lmbU gene; lmbV gene; lmbW gene; lmbX gene; lmbY gene; lmbZ gene; lmbA gene; lmbR gene; lmbC gene; ORF1; ORF2.
SOURCE Streptomyces lincolnensis
ORGANISM Streptomyces lincolnensis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
REFERENCE 1
AUTHORS Peschke,U., Schmidt,H., Zhang,H.Z. and Piepersberg,W.
TITLE Molecular characterization of the lincomycin-production gene cluster of Streptomyces lincolnensis 78-11
JOURNAL Mol. Microbiol. 16 (6), 1137-1156 (1995)
MEDLINE 96020646
PubMed 8577249
2
REFERENCE 2
AUTHORS Piepersberg,W.
TITLE Direct Submission
JOURNAL Submitted (10-MAY-1994) W. Piepersberg, Bergische Univ. Geamtschule Wuppertal, Inst fuer Chemische Mikrobiologie/FB9, Gausstr. 20, 42097 Wuppertal, FRG
REMARK 3 (bases 1 to 36270)
REFERENCE 3
AUTHORS Piepersberg,W.
TITLE Direct Submission
JOURNAL Submitted (13-JUN-1994) W. Piepersberg, Bergische Univ. Geamtschule Wuppertal, Inst fuer Chemische Mikrobiologie/FB9, Gausstr. 20, 42097 Wuppertal, FRG
COMMENT On Jun 14, 1994 this sequence version replaced gi:487688.
FEATURES
source location/Qualifiers
1. 36270
/organism="Streptomyces lincolnensis"
/mol_type="genomic DNA"
/strain="78-11"
/db_xref="taxon:1915"
1904..3363
/gene="lmbA"
1904..1907
/gene="lmbR"
1918..3363
/gene="lmbA"
/note="bases 318..1763 in X59926"
/codon_start=1
/transl_table=1
/protein_id="CAA5745.1"
/db_xref="GI:487689"
/db_xref="GOA:P46104"
/db_xref="UniProt/Swiss-Prot:P46104"
/translation="MSVFAKTSLSFRAKRADEARSRKVTLVFLAVQLLLA
VDVTVNIALPARDSPHVDTRQLTWVTGTVGGGLMWGRIADLFGRRITLFG
AFLFGASSLAGLAPNDELVLARFGQAGALSLPAMSLIACSSRTAPQGVRLA
SVAAGVIGLGLGVTQLPSKRMIFLINPLVSLVAVLLVKDEETARVDL
FGALFTPAAPLLIFGVNELGEDPRLPLAVGSLAAVCAAFVAVERRAHLVPL
TFPCNRVIVANGATVLSALSTSPLLTMHLOBERDISPIEAGLSFLIGSLIIA
CVLRGILERTGTGAATVGMALGPRRLPALPSDNLSTVSFPGKILLRATGL
VALONALHAATEADAGVAGVQRCADQLGASGIAVYVSIIGFSPHGGWDPTVAV
SLAIGIGIAVLAVALVLSPRRLAAPROBD"

gene /db_xref="GI:487690"
/db_xref="GOA:O54353"
/db_xref="UniProt/TxEMBL:O54353"
/translation="MLPSKEPLSGTTGAVSATHWLASNAGKRILBENGNAADAVAAG
FLVQVBPBPNRPGGDVDSIGRTARRADDAEILCQGPMPRATPOAFSDLEHILPGS
GLPACVPGAFGWLRLAEFTLRADVLAATVYBEGPLPETAHADVLAFLP
RTWETGSGQIYLIANGAPKAGSRFRNPAQVYQLKEAASADREAOIEAAR
YOGCAAVORFLDGGRLDAGSALLITADLAATSBVTTRATSRPGAVYVHK
GPGSGVPLQQLALLEGFDLAGRPDSADYHIVCEAKTAPADRAWYDDPFTDI
PLADLSDDYTRERRLVGPBAAVNELOPGTPGSSWLPGAAEPBEDLSGIDEMWG
QLRNLGPTILKATTAKGDTCCVTVDRHGNTVAATASGMLKSSPALAEIGPLGTG
QTVYIAGHPNLSIAGKRPRTTSLPTVVGSDGPHLAFTGPGDRDOMTLQFGLVY
AFGLDQATETLAFHTDOVPAFTSPHOSRPGVVVEBNCAATVAELTRGHRERY
PAYSIGVCATGCTDGLVRAACPRRQPAIGC"
join(5618..5622,5628..6104)
/gene="lmbB1"
5618..5622
/gene="lmbB1"
5628..6104
/gene="lmbB1"
/codon_start=1
/transl_table=1
/protein_id="CAA5747.1"
/db_xref="GI:487691"
/db_xref="UniProt/TxEMBL:O54354"
/translation="MPSVKGMPVSVHVGVOQTDLNLSIKNQEPFGCVSWTLDF
SALHRSPLGIERLAERIGDYRPHHIGVSGAERPAEANOQHVCFAVGSFTELB
AMRSRWLELYARGWTFAPVERATDIDVDKGVNSFYALDPNGLEBFTYVDPGR"
6101..7054
/gene="lmbB2"
6101..7054
/gene="lmbB2"
/codon_start=1
/transl_table=1
/protein_id="CAA5748.1"
/db_xref="GI:487692"
/db_xref="UniProt/TxEMBL:O54355"
/translation="MSSLFARRTDRTDPLPAAQDWYGVGYVGLPPLTLPYAPSGS
PAARRRGRRPALAGHLADPVAGVPANNAVDLTDPGVDRLPFRNITGQVFLM
QLASVLAESAEGGEGARAERARRVRGSLMLTYSQPSRVYDRLLRPHALQ
RHLSGAARDYHPVRSLLRGLPAKGLDAPLRECRYNHHHEGIAKLVSQVSLQ
QTNHOREFLHRDLRLSLYDCVFLTVRAPASVQVNTQVLRHAIGQDLAANGLYP
AYAPSGHEBPAAEPDVARCKETLPLDLIRICDSATVAAAS"
complement(7174..8700)
/gene="lmbC"
complement(7174..8700)
/gene="lmbC"
/codon_start=1
/transl_table=1
/protein_id="CAA5749.1"
/db_xref="GI:487693"
/db_xref="GOA:O54356"
/db_xref="UniProt/TxEMBL:O54356"
/translation="MSSSVRLSTLAGAAEREPAPAVLGETRLPYVYELARAGGI
QALLAGVRAQGRVALISGRTDADAAYAHALIAAVVPLDAAAPPARMAVSRY
CAPYAVGERALIDRPAAYGRRRLALPSDGBELPAGRLDAYASDSTDVAYLLTSSG
STGVKCVATITGAAPSPMSVMGAFVGGDDVPACAPLHFDVSVASVLSAGLGGAA
LAPVPRLSGFPVELATWIAERATVWLSPYPLATVSGIEERARERLARLRTVFA
GDVFPORLALICAPARPLNTYGPETENGCTEVVDADPAGVPVIGRPEVCAEW
EDDGRVVDVAGSGELVAGPTVAAAGVGGHGAFRGETCPGSAAYRTGDQVY
VLGGRGAFVAGSDNMIMKMGORELEBVENAVLADGVECCCVKQVDVDRHSRLLA
VVVPGAGDRTLRLEHCTLRPLPSNAVYHRLTAAALPLSGTKVDRARALRELYTRGE"
8732..9787
/gene="lmbD"
8732..9787
/gene="lmbD"
/codon_start=1
/transl_table=1
/protein_id="CAA5750.1"
/db_xref="GI:487694"
/db_xref="UniProt/TxEMBL:O54357"
/translation="MAVSPQSVRLPEHAPHCEIGCANVAKRVGVDFPRVYAGSQWLA

```
HLRRERPGDIPOPMIDRTHLTREGGTARAGPPLAGITDGTVEHPVPAARSELPVLRNDLIA
RGGSLIFLEDRBCFWLHSGPGLPHTVTPGVDAAGWQIIEGHSWAGAPMAEAD
LLAAAYRPDDPHNAGRVSLTALTPERRAALDALDRDSVAYLTLGGGRGLDTPA
GTLVATDGGAAVAELDLRLRGWEYLCAALADGAVEGIDIAVGRVLTFLGLADELAY
TSYARTGARLRLRGLALADIGDEHLPIVWQORAMRTPSGSTRCASTDALHLHDID
AAGVADACARLAEVL"
gene
complement (9805) . 10629)
/gene="1mbx"
complement (9805) . 10617)
/gene="1mbx"
/codon_start=1
/transl_table=11
/protein_id="CAAS5751.1"
/db_xref="GI:487695"
/db_xref="UniProt/TREMBL:Q54358"
/translacion="MTQCLITVHAHPDDPDESRGATVAYHTAOCVAVLVTCTDGGAG
EVINDAVTDPTTPRPFVAVRBAELDASARNLIGSAVRILGYSRSGDGTAGAEAVR
APLDEAARLRLVADERPDVIGYGNHTRDPHPIRANVELTRERVLDDHTPAVY
HIAFSRRHRALHQAQVDSGVSPYEGSLAPGAPGAPDEWITTLVDTGDAVERLID
ALRSHVTVQVPASGMFALSPQLRDAPPEYETRVGAAPREAVVHDLFTAPA"
RBS
complement (10625) . 10629)
/gene="1mbx"
complement (10665) . 11945)
/gene="1mbp"
complement (10665) . 11945)
/gene="1mbp"
/codon_start=1
/transl_table=11
/protein_id="CAAS5752.1"
/db_xref="GI:487696"
/db_xref="QGA:Q54359"
/db_xref="UniProt/TREMBL:Q54359"
/translacion="MTATASGATAPAEITDGMILIRTSVDYVPRPDDPSAOMKLD
RAQARLPVLYLHVADAEATPGLRBAALADAPCDGLLFSQYGLPELRRLDMLTA
ADEEMPTAEPPLVSVAMSGAIFDPRMLKDGSPSAVILPRPGCYMSYRDTG
HVEVSVEPSPHGDPAHLEAORCSBGLDVACILINQHPKNGANTPEFLAA
VALAERERPVLDNAFVGLTAEVDRPSAVRILGHLVGLSLVSVLSKOPACSG
MALGAVAGSLVAVSAGRWRCLEPRTAFRAQAAALMAGAEERFTRRREBART
HARLRTTRGGLPDPAVTHHGAAPPTTLRPPGSGTVEBEVQTVVRHGVLLGERDA
RRRPFWKVLGSDSVFEPARALGDAAEMWYR"
gene
complement (11942) . 12742,12748) . 12753)
/gene="1mbx"
complement (11942) . 12742)
/gene="1mbx"
/codon_start=1
/transl_table=11
/protein_id="CAAS5753.1"
/db_xref="GI:487697"
/db_xref="UniProt/TREMBL:Q54360"
/translacion="MRDYRLMHTTAAQVSPPLVWMLQVTRGATVADVCGGAGNF
LYRAMGHTESWLKQITRPGILGLDMSPHATPLAERHRYDRHLCPATLTLDD
VVDLTCETTELHYVPRVAVRELRANVATQIVITTPVEHIVNPELLTETIDAR
RQDDMPREYAVVLESALHAKMLSTERLAAGFTDAVVGSRILYHADPAKMDPEAL
GEVPGIPIPPAREBEADMRERYVQLLESVREMAVAPGLTAGATR"
```

Alignment Scores:

```
Pred. No.: 2.7e+03 Length: 36270
Score: 110.50 Matches: 65
Percent Similarity: 30.00% Conservative: 25
Best Local Similarity: 21.67% Mismatches: 103
Query Match: 7.56% Indels: 107
DB: 1 Gaps: 13
```

US-09-087-136-1 (1-275) x SLILNC (1-36270)

```
Qy 21 ArgInAsnAepProLeuIlleSerGlyGlyProLeuProLeuGluSerProAsnArglys 40
Db 21638 CGGACACGCGTCCACCTTCAGCCGCTGACCCGCTTACCTGCTCGGACACCGCTCCT 21579
Qy 41 LeuThrSerLeuLeuSerTyrAspProThrValProGluSerProAsnMetLysPheAla 60
Db 21578 -----CCGACCGGCGCCAGGCGTTCGCCGCTCC 21552
```

```
Qy 61 ArgLysArgLeuGlyAsnLeuLeuThrThrIleLysHisHisProSerGluIleIleGly 80
Db 21551 CAGCGCGCGCGGACAC-----GCTCGCTCGCTGCC 21519
Qy 81 ValLeuProGluAAspTyrThrArgAlaAspGluGluProGluArgGluGluArgProPro 100
Db 21518 CGCTTCCCCCGC---CAGCAGCAGCGCGCGGTCTCCCGGTCTGGCGCCCGCTCGCGC 21462
Qy 101 GlyArgProArgLys-----MetProArgHisGluLeuSerThrSerLeuMet 116
Db 21461 GGC---CCAGCAGCAGCGCGGACCGGTGAGCGCCGATGTCGCGCGACCGAACAC 21405
Qy 117 GluSerProArgLysThrMetThrArgAspSerLysIleMetThrGluLeuArgGlyLys 136
Db 21404 GATGACCTCTGCTCTCTCTCGCGCGCGC-----GGCGGCGCAC 21369
Qy 137 ProThrGluMetIleAlaGlyArgPheGluGluGluTyrSerLeuGluArgAlaTTPVal 156
Db 21368 CCGCAT-----AspTyrLeuAlaCysArgGluIleHisArgMetPro 192
Qy 157 LysGlyHisMetAsnAsnGluTyrGluProIleLysAlaGluArgThrAspTyrAlaPro 176
Db 21356 CGTACGGCGCAGCAGCGCGGACGCTCGCGGTGACGCGGACCTTCCTCGCTCG 21297
Qy 177 AsnLeuAlaVal-----AspTyrLeuAlaCysArgGluIleHisArgMetPro 192
Db 21296 GCGCGCTCGCGCTGACCGGCGGCGCGCGCGCAGCATGTCGTGTGTGACCGGACCA 21237
Qy 192 ----- 192
Db 21236 CGCGCGCAGGTCCAGAGAGTGATCGGGAGCGGTCGCCGCTTCGCGGTACCCA 21177
Qy 193 -----ArgProAspLysSerIleProGluLeuProIleValPro 205
Db 21176 CAGCGCGGAGCCCGACCGCGCTCGCGCGATCGGGCGCTCGCGCACGCTTCGCT 21117
Qy 206 SerArgIleAspGluPheAspAlaThrValAsp-----ProArgTyrGlu 220
Db 21116 CCGCGA-----GACCTACCTGCTGACGATGTCGATGCTGACCGACCATGCA 21069
Qy 221 ThrAspLeuLysAsn-----GluTyrIleArg 229
Db 21068 CGTGCCTCTCTTACCGCGCGCGCGGAGCGCAGCAGCATGCCGCGGATGCGCTCG 21009
Qy 230 HistTrpLysGluValLysLysGlyTTrpCysAlaHisGlnArg----- 243
Db 21008 CAGCGCGCGCAGCGCGCGCGGACCAATGCTCTGTCGACGCGGACGAGTGTCTCTA 20949
Qy 244 ArgArgThrAlaProHisAlaArgSerIleAlaLeuIleAsnLysIleTyrGlnProGly 263
Db 20948 CGCCAAGTCCCGCGCGCGCGCGCGCGGACGACGATGCGGACCGGACTGCTCGCTCGA 20889
RESULT 13
AY498874
LOCUS AY498874 41987 bp DNA linear BCT 08-JUN-2004
DEFINITION Streptomyces lavendulae linear plasmid pSLV45 SLV.1 (SLV.1), SLV.1B (SLV.1B), SLV.2 (SLV.2), SLV.3 (SLV.3), SLV.4 (SLV.4), SLV.5 (SLV.5), SLV.6 (SLV.6), SLV.7 (SLV.7), SLV.8 (SLV.8), SLV.9 (SLV.9), SLV.10 (SLV.10), SLV.11 (SLV.11), SLV.12 (SLV.12), SLV.13 (SLV.13), SLV.14 (SLV.14), SLV.15 (SLV.15), SLV.16 (SLV.16), SLV.17 (SLV.17), SLV.18 (SLV.18), RegA (RegA), TrsSLV4 (TrsSLV4), SLV.21 (SLV.21), ParsSLV (ParsSLV), SLV.23 (SLV.23), SLV.24 (SLV.24), TpsSLV (TpsSLV), SLV.26 (SLV.26), SLV.27 (SLV.27), TrsSLV2 (TrsSLV2), SLV.29 (SLV.29), SLV.30 (SLV.30), TrsSLV, SLV.32 (SLV.32), SLV.33 (SLV.33), SLV.34 (SLV.34), SLV.35 (SLV.35), SLV.36 (SLV.36), and SLV.37 (SLV.37) genes, complete cds.
ACCESSION AY498874
VERSION AY498874.1 GI:42794855
KEYWORDS Streptomyces lavendulae
SOURCE Streptomyces lavendulae
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.
```

REFERENCE	1 (bases 1 to 41987)
AUTHORS	Hosted,T.J., Wang,T. and Horan,A.C.
TITLE	Characterization of the Streptomyces lavendulae IMRU 3455 linear plasmid pSLV45
JOURNAL	Microbiology (Reading, Engl.) 150 (Pt 6), 1819-1827 (2004)
PUBMED	15184568
REFERENCE	2 (bases 1 to 41987)
AUTHORS	Hosted,T.J., Horan,A.C. and Wang,T.
TITLE	Direct Submission
JOURNAL	Submitted (10-Dec-2003) Functional Genomics, Schering-Plough, 2015
FEATURES	Location/Qualifiers
Source	1. .41987
	/organism="Streptomyces lavendulae"
	/mol_type="genomic DNA"
	/strain="IMRU3455"
	/db_xref="taxon:1914"
	/plasmid="linear plasmid pSLV45"
	complement(1803. .2057)
gene	/gene="SLV.1"
CDS	complement(1803. .2057)
	/gene="SLV.1"
	/codon_start=1
	/transl_table=11
	/product="SLV.1"
	/protein_id="AAS45783.1"
	/db_xref="GI:42794856"
	/translation="MYGAKPSRGPGPKGHVYVYLVGPIDGMLDVTGMDSGEYVD
	GAMLSHGARGPGRSDYEPADPGAGVGRFVWRGDVP"
	complement(2343. .2579)
gene	/gene="SLV.1B"
CDS	complement(2343. .2579)
	/gene="SLV.1B"
	/codon_start=1
	/transl_table=11
	/product="SLV.1B"
	/protein_id="AAS45784.1"
	/db_xref="GI:42794857"
	/translation="MGESGSGPAGRGGRORRRMAAVRVAGRVAEGLAVEAVRAAIV
	SVMAPEMSDVPPLETAVQALRLRRGGRPRG"
	3104. .3475
gene	/gene="SLV.2"
CDS	3104. .3475
	/gene="SLV.2"
	/codon_start=1
	/transl_table=11
	/product="SLV.2"
	/protein_id="AAS45785.1"
	/db_xref="GI:42794858"
	/translation="MOGRAGAAGVGLNVSVPDLPRLTLVTYLRGELSRTEQALAA
	AEEERATAROPAPERPAMLYVERGIGVRLPVRVHTGDCMDTGKRCVPAPADQVRL
	LAEGVACIHRCPDIALGVLE"
	3499. .3723
gene	/gene="SLV.3"
CDS	3499. .3723
	/gene="SLV.3"
	/codon_start=1
	/transl_table=11
	/product="SLV.3"
	/protein_id="AAS45786.1"
	/db_xref="GI:42794859"
	/translation="MSPDLRVITYPPDAQGRVRVVDGEILGRALGPADLLEFRRAGS"
	LDPPDVNDLPLLEBMRGCPAVVTPDASG"
	complement(3737. .4069)
gene	/gene="SLV.4"
CDS	complement(3737. .4069)
	/gene="SLV.4"
	/codon_start=1
	/transl_table=11
	/product="SLV.4"
	/protein_id="AAS45787.1"
	/db_xref="GI:42794860"
	/translation="MLAAGVAEYLMTHEQGPVADAERMALEAGGIDGGLARDLVDS"

gene
LAPVMEPLDVGPEGPAPRSMMWHQROKQATTYERWMPQLEAVADLVCHRLNPAS
ILEVLTQK"
4362.5435
/gene="SLV.5"
4362.5435
/gene="SLV.5"
/codon_start=1
/transl_table=1
/product="SLV.5"
/protein_id="AAS45788.1"
/db_xref="GI:42794861"
/translation="VTRTPSPAADROLIQACHREPIDKASRLDAMPRHGILSPSEV
PFGKRGKQRYPEGTLEQVCLAVCDOLQPKSPGDLILTFPAENPLPLPGPK
CALAVYEGAPFQAPAEQARAEEVPDLEBELPADFNMAETAQLERDPEHVRQM
RENILRLDLRAGROELANVNGILTNALNRLPVEDTYLMDLKKAMFDGPAPAS
RAMAHALCGEQLAVRFTNATERFEGILATGETILHRELVLANMAAYVASE
GLLSVEYERLSSAATLACGMLVENMSKQVPPSGSRAQLAFELRSLWVCSHSGGG
GEGRRAPAAIRISGCVTPRRPCR"
5680.8124
/gene="SLV.6"
5680.8124
/gene="SLV.6"
/codon_start=1
/transl_table=1
/product="SLV.6"
/protein_id="AAS45789.1"
/db_xref="GI:42794862"
/translation="MGSTPESSHLGDPQIPQISOIAELAAVERPVSNMKRLSDPPT
PRTAGTTPYLRSDILRPLDRILPSRRRLPGEGALTYGARARSILAMPKQGRSA
RPAALSVLMSSTDEEPDQGREVYKLMPLADLRIGSGSLENTYHLVALCLPRA
LSHDWEVTRVAGVAAAAGRGSAAQTGAGRGVGEVNLIGRTVDNELQKQKORSVTA
TSLSNLTPRRAGDLALVMNLVNOIGNAPFLUIDAYEKHATLRSEFTPDGVVRMAVA
SLARSTVGRPHVTYDPIYRGGEFLAESVTDGASILNSDEPLAVTGERTDDPPLAL
LAGNLVLGGRPRRLVYHKAPMAEIRGGEPAADLVLTNPRFMKQAGACGEGTM
AYGAPVNDMLAYQHALSLRAGGRALVMPFKAGNSAAETAIRMAVQAGVBE
CTIAPALPVGSTAVPVSWLIRHPDCEVFLPDAHLCVGRPPCYLKEDVQVQA
LGTAYPALVSGTAVPYENPQPLHVPSSAVERQOILDRACSLNPVDHISAQSAATGEGAG
SVDEASGVQDQIELKESINRADAHVAFLSQVGENGEGAGGGENLAEILCKQKAG
PFTFGVAGKTPNGPVLVMPRHLKNGNITDTPERSQELADRHHRHLYLQDQICLQ
VRSQGVTPALVADQSGMLMSTVYIRLRVHEGDEVSNYTFRLGPESLAMTVDSRS
AATAPSSITKTLGTMTRYRLPPLPQORIALDLRDEQRAHNHNLAEISRSLSLA
EOLMAPGSRSLTSLGFS"
8147.8473
/codon_start=1
/transl_table=1
/product="SLV.7"
/protein_id="AAS45790.1"
/db_xref="GI:42794863"
/translation="MISRSRQRCRSGMACPSNHPKQTHRVMTATDIPDVTVSRTM
PWPMLTLMTFTFTVGMASGLVESLASAGLPVIRAVGTGFLSAGLCTIAPVVIL
QLRKRS"
8909.9136
/codon_start=1
/transl_table=1
/product="SLV.8"
/protein_id="AAS45791.1"
/db_xref="GI:42794864"
/translation="MSPLRIYIVYPRAGEGRARVVDGEILGRALGNDLFEFLRARGAG
LDPDVVDLADELFTIRMGSGGPAVMSNGDTE"
9288.9626
/gene="SLV.9"
9288.9626
/gene="SLV.9"
/codon_start=1
/transl_table=1
/product="SLV.9"
/protein_id="AAS45792.1"
/db_xref="GI:42794865"
/translation="MLATGVEAEILIBRQGSVTDABRWAVEAGLDLDOGLARVVDG
LRYHGWPLIDGRLGAAAPHSMWPHQKQKQATTALEHNMPOLEAVANLLVMRRLLPAD
ILVLEBQSG"
9970.10563
/gene="SLV.10"
9970.10563
/gene="SLV.10"

```

CDS      complement (9970. .10563)
         /gene="SLV.10"
         /codon_start=1
         /transl_table=11
         /product="SLV.10"
         /protein_id="AAS45793.1"
         /db_xref="GI:42794866"
         /translation="MDGQGNLTTLTPHPPLPPPPPPRAHPAPPAAPPAAPROQNT
         VOMGAPIPEPTHONAMAOIMAAHAAADLTAVAGAKRIEALAEVGEPLHPTTALGA
         AVELSGPLVDMLTRKWRKESRAQDVMRWADNSMRPY"
gene      complement (10802. .11983)
         /gene="SLV.11"
         /complement (10802. .11983)
         /gene="SLV.11"
         /codon_start=1
         /transl_table=11
         /product="SLV.11"
         /protein_id="AAS45794.1"
         /db_xref="GI:42794867"
         /translation="MEGRDALRAAASVAGRAALKYGATGSAVLILFLIVITGMSS
         GSKALADACGGRWGPELPSDIDGSAQOLPGSTRKDQIKNAQIIDETAKRGISGR
         ATLALMTALQESTLTLNLDHDDUSIGLFOQPSGSGVTAQIMQPKVATWFFRGD
         SGDPBGJTDIRGEMETPLGAQAQVQSAIPDYDHEHGPAREIATGAIIDINRAGTS
         TGNRPGRTGTQGPRTNRCDTNGTPGQGPFPDHPGQAPPMKYNINPSTGATLAWA
         KAEAKNGSEWTRACIAFVARTYGMGSAVSFAIDHYIOWMPPEMKHDKRNPAGALM
         YWETGSRAGHVALYLGGMIAVANDIKRPYIDIVPATDIETKMGATYVGMAPPYRPG
         G"
gene      complement (11987. .12472)
         /gene="SLV.12"
         /complement (11987. .12472)
         /gene="SLV.12"
         /codon_start=1
         /transl_table=11
         /product="SLV.12"

Alignment Scores:
Pred. No.:      3.2e+03      Length:      41987
Score:          110.50      Matches:      67
Percent Similarity: 35.56%      Conservative: 34
Best Local Similarity: 23.59%      Mismatches: 90
Query Match:     7.56%      Indels:      93
DB:              1          Gaps:      16

US-09-087-136-1 (1-275) x AY498874 (1-41987)

QY      15 AspGlyAspArgAsnAlaArgLysAsnAspProLeuIleSerGlyGlyProLeuPro--- 33
DB      39995 GATGGGGAGCAGACGGCC-----GCCTACTTCACCGGGGGCCCTGTGCGCGCC 40042

QY      34 LeuGluSerProSer-----ArgGlyLeuThrSerLeuLeu 45
DB      40043 CTGCGCGGGCCCGGAGCGCCAGGGGATCTGCTGCAGCCGCCAGCGCGCGGCTGACTG 40102

QY      46 SerTYrAspProThrValProGluSerProAspMetLysPheAlaArgLysArgLeuGly 65
DB      40103 AAT-----GTGCTCTCCAGAGCTGGGAGCACTCAAGATCCGACCCGCTCTCT 40153

QY      66 AsnLeuLeuThrThrIle-----LysHisIleProSerGluIleIleGlyValLeu 82
DB      40154 CCGGACCTCGTGTATATCGCGGGGTGGAGCATGCGCGCGGACCGCTGTGCAAGCTTC 40213

QY      83 ProGluAspTYrThrArgAlaAspGluGluProGlyArgGln-----GlyArgPro 99
DB      40214 -----CGGAGACAGCGCGCGCGGCAAGACGCGCGGACGCGGAGCGCGG 40255

QY      100 ProGlyArgProArgLysMetProArg-----HisGluSerSerThrSerLeuMet 116
DB      40256 AAGGGAGAGCGCGGACCGCGTCCCGCGGTGAGAGTGCATGCGAGGTGCGCGGCTCTC 40315

QY      117 GluSer---ProArgLysThrMet-----Thy 124
DB      40316 GAAGCGGACCGCGCGTCAAGATCGCGCGTGTGCGCGGAACTCGGCGTGGCGCGCGCC 40375

```

```

QY      125 ArgAspSerLysIleMetPheGluLeuArgGlyLysProPheGluMetIleAlaGlyArg 144
DB      40376 ACGGGCAGCGGGCCCTGTCTCAGCTGCGGACGCGAGCAATCGAGCTTTCATGCGC--- 40432

QY      145 PheGluGluGluTYrSerLeuGlyArgAlaTrpVallyseGlyHisMetAsnAngluTYr 164
DB      40432 -----

QY      165 GluProIleLysAlaGlnArgThrAspTYrAlaProAsnLeuAlaValAspTYrLeuAla 184
DB      40433 -----GGACAG-----CCGTTTTCAGCTGACGACGAGCGCGG 40465

QY      185 CysArgGluIleHisArgMetPro-----ArgProAspLysSerile 198
DB      40466 GCCCGGGGT-----CGGGTACCCGGCGCGGTCCGGGGCGGACCCGCGAGCGCTTG 40519

QY      199 ProGluLeuProIleValProSerArgIle-----AspGluPhe 211
DB      40520 TGACCACTACCAACGACGACCGGACCGGAGAGCCATGCTGAGGCGCTTG 40579

QY      212 AspAlaIleValAspProArgTYrGlyThrAspLeuLysAngluTYrIleArgHisTrp 231
DB      40580 GACGTGAGTGTGTGGAGACGCTTACTACGCCGAC-----GTCGCCACCTTC 40624

QY      232 LysGluValLysLysGlyTYrCysAlaHisGlnArgArgArgThrAlaProHisAlaArg 251
DB      40625 CGCCCGGTGTGAGACGACGAGAGCGGAGAGCGCTTCCGCGATGCGCCCGCCGACCGGGA 40684

QY      252 SerIleAlaLeu 255
DB      40685 CTCTCGGCGCTC 40696

RESULT 14
STU82965/c 69301 bp DNA linear BCT 19-JUL-2002
LOCUS STU82965/c
DEFINITION Streptomyces toyocaensis strain NRRL 15009 biosynthetic gene
cluster A47934, complete sequence.
VERSION U82965 Af039028
KEYWORDS U82965.2 Gi:21911402
SOURCE Streptomyces toyocaensis
ORGANISM Streptomyces toyocaensis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomycetes.
REFERENCE
AUTHORS Marshall,C.G., Broadhead,G., Leekiw,B.K. and Wright,G.D.
TITLE D-Ala-D-Ala ligases from glycopeptide antibiotic-producing
organisms are highly homologous to the enterococcal
Vancomycin-resistance ligases VanA and VanG
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 94 (12), 6480-6483 (1997)
MEDLINE 97322398
PUBMED 9177243
REFERENCE
AUTHORS 2 (bases 1 to 69301)
TITLE Pootoolal,J., Thomas,M.G., Marshall,C.G., Neu,J.M., Hubbard,B.K.,
Walsh,C.T. and Wright,G.D.
TITLE Assembling the glycopeptide antibiotic scaffold: The biosynthesis
of A47934 from Streptomyces toyocaensis NRRL15009
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (13), 8962-8967 (2002)
MEDLINE 22080153
PUBMED 12060705
REFERENCE
AUTHORS 3 (bases 67313 to 69301)
TITLE Marshall,C.G., Lescaud,I.A.D., Park,I.-S. and Wright,G.D.
TITLE The origin of vancomycin resistance?
JOURNAL Unpublished
AUTHORS 4 (bases 1 to 69301)
TITLE Marshall,C.G., Leekiw,B.K. and Wright,G.D.
TITLE Direct Submission
JOURNAL Submitted (20-DEC-1996) Biochemistry, McMaster University, 1200
Main St. W., Hamilton, ON L8N 3Z5, Canada
AUTHORS 5 (bases 63100 to 67312)
TITLE Marshall,C.G., Leekiw,B.K. and Wright,G.D.
TITLE Direct Submission

```

JOURNAL Submitted (20-DEC-1996) Biochemistry, McMaster University, 1200
REFERENCE Main St. W., Hamilton, ON L8N 3Z5, Canada
AUTHORS 6 (bases 67313 to 69301)
JOURNAL Marshall, C.G., Lessard, I.A.D., Park, I.-S. and Wright, G.D.
TITLE Direct Submission
REFERENCE Submitted (11-DEC-1997) Biochemistry, McMaster University, 1200
AUTHORS 7 (bases 1 to 69301)
JOURNAL Marshall, C.G., Leskiw, B.K. and Wright, G.D.
TITLE Direct Submission
REMARK Submitted (16-JUN-1998) Biochemistry, McMaster University, 1200
REFERENCE Main St. W., Hamilton, ON L8N 3Z5, Canada
AUTHORS Sequence update by submitter
8 (bases 1 to 69301)
JOURNAL Pootoolal, J., Thomas, M.G., Marshall, C.G., Neu, J.M., Hubbard, B.K.,
Walsch, C.T. and Wright, G.D.
TITLE Direct Submission
COMMENT Submitted (11-JUL-2002) Antimicrobial Research Center, Department
of Biochemistry, McMaster University, 1200 Main Street West,
Hamilton, Ontario L8N 2S, Canada
On or before Jul 19, 2002 this sequence version replaced
gi.4104707, gi.3228363.
FEATURES
source Location/Qualifiers
1. .69301
/organism="Streptomyces toyocaensis"
/mol_type="genomic DNA"
/strain="NRRL 15009"
/db_xref="taxon:55952"
complement (766. .2022)
/note="Integral membrane ion transporter"
/codon_start=1
/transl_table=1
/product="Stac"
/protein_id="AAM80527.1"
/db_xref="GI:21911404"
/translation="MLAGLAVTFAVAFVAFALAKRIQRPVWGEIIGALAPSLGL
LPGPVEALFPADRPVLAQSOLGVLFMGIGELDFALLRGSGRVTLVSMSA
LPMALGAGLAVLPVLDKQSLSDGVLAAPLIPGVAMSTAPAPLARIATENMORT
RIGMALACAVQDFLAKVCUAVAVVITASGLMPRLAMVIESGLFLGVLLIVRPGI
RRLAPERRRAVGSLLIHAILVTGLVSAWMTHEIGLHVGAFAGFVAVPHRIEVO
APQVERIEQTSLFLPVFTVTGLSVVGGIGRTGLMTAVLVAVCTGCVAAVAS
ARLTGASAEERTLTGLNARGLTVELVTLVINGLGLVLSRLFSAMVMALVTTFMG
PLHRIQHEELRIIVPAVKNADNSV"
complement (2485. .4071)
/note="putative nonheme iron dioxygenase"
/codon_start=1
/transl_table=1
/product="Stac"
/protein_id="AAM80528.1"
/db_xref="GI:21911405"
/translation="WTEQLFLRSDDTIIEPLVDRPVAQLYTTAVSGTMNIAFQIEM
LBSYLQTPDHFHSGRDPKFRGFVGLIEERKEKEVEALAAIRDRDMKRFEAIA
EAEKIVLEATGYDLRLSYPLPSBELAGLVVYDTGQASLHLEPLAYQSKATED
ROVSLEETGERPFWMTSPRLPSDVLIEDIPRHGIBELFLSRIRPTVALE
ALIEGDAEAQLSLVLAEBGLAADRYTAGRTIRYMGHACILLQTPVSTLTPPTS
ADRTATRYTDDLPDHDIVYITTHGSDHLYATLIDLRGVGTFFVPTSRKSLCD
PSLALFKICLGLPVI EVDVDFEVPFGGKIATPFGHADLDIRAKSTWISIGGS
VYVAGDSGIDPVLRYVYRRLHGVDI AFLGMECDGAPMMKXQPMLTQPIPKMSDT
RKLGSNAEQSAIVTELGAEBYVIYAMGESMWLGHYMATSYNEDSYQLQIABFEEM
CTRGQVKAHGLNRAEMHWA"
complement (4151. .4963)
/note="bulfoctanferase"
/codon_start=1
/transl_table=1
/product="Stac"
/protein_id="AAM80529.1"
/db_xref="GI:21911406"
/translation="MNGKCYLASYPKAGSHLRCMLTSYTGGEVETWPGIQAAPVHL
EGLRDEGASPADPEOVLLATHTFADRPVLRFPRESTAKVCLIRNFRDMLSLMNR
KGIPEDEACRKIAETFIADGESSVAVIMAGBSWENISWDSVHESFPNAAVLA
VRYEDLRDPEGLMKVDFLELGGDVAIVANCTLERMRMEERSKLGLGTTG
MTRGKOLPFGVGGORSKLTKFMGDIEKAVADLLHGTBDPAHYALYGAH"
complement (4960. .6267)

/note="Nonheme halogenase"
/codon_start=1
/transl_table=1
/product="Stac"
/protein_id="AAM80530.1"
/db_xref="GI:21911407"
/translation="MSAETFDVVVAVGAGGAPAGSTAASIVAMOGHRLVLEKEVLPRHRV
GESLPSITVHGVCMILGVTDEIARTDPVVKRGATFONGRAEPMAFSGASRLDAT
SPAYOADRAPDEILIKNAKRGVVRGSGSVTA VVEBGMGTGRLRHTDPDGGHEVG
ARVYTDAGHESPLSASVGTGRDPAAGFRGLALSQVYRNGCKRPLPERSGGSLSTPD
GGMFWYVPLFGGLTSVGAIVVSEEAERLQDREAFALINCEPAISDLSSEGMKVA
GKXGLVREDDYVYHRTTFMRPGMLLVGDAAACVDPPLFSSGCVHLATYALSALAAASINS
VLADGVDKXTALTEFEARYRRELGA FSELLITTFPTMGEEHPESFVDVAGVSSGE
WALAAGGAAAPPEMPFRGLAVSPGCMKMLPAASEREDAR"
complement (6312. .7490)
/note="P450-related oxidase"
/codon_start=1
/transl_table=1
/product="Stac"
/protein_id="AAM80531.1"
/db_xref="GI:21911408"
/translation="WRRTLCDPHEDMPLRAHGLIRIIEGNASQWSTDYVQWAGVD
VVRKILGDHENFTRLRLTDAQPLSGGVSVPPELAQISLYDPEHTRLRLMLTPEP
TVRRIRLEPAIEGIIIEHLDALGAGPPADQLVLFADPVGERTCEILGVPRDRNE
FIRRVQNVDSRGYKARADSAFNRYLMTLIRORNDPESGRLMVLVREHGRITD
BELKGVCTTALIGGESVAGMIGRGVALLLEHPORRLPESGRLMVLVREHGRITD
AVOQPTPRMAVRDVPVREGQLIKAEHYTLCSLIMANRRBGLTSDSHLDANREPLPHVA
FGHGHICGAAVAAVRLITYQSLMRFPRLSLAVPAGEVGFNRNAFLTDSFRLPVTW
"
complement (7534. .9072)
/note="Nonheme halogenase"
/codon_start=1
/transl_table=1
/product="Stac"
/protein_id="AAM80532.1"
/db_xref="GI:21911409"
/translation="IMDGLGDPPLSTIAKXISBGWEMSVETPDUVVAAGGPGGSTVA
ALVAMOGHRLVLEKEVPRYQIGESLIPSTVHVCPLLGVTDLSLAAPKPRVGGTF
RMGARPEPMTFSFVSPRMTGPTSFAYQVERARDEILINNAKSKGYTVREGCYTA
VEBGRDTGARLYTSDSGEHRVGA FVVIDSGNKSRLHAAVGGTRESSEFPRSLALG
YFENGKPLAEYSGNLSIVAFDSGMFWYIPLSDTLGVGVREMAEKLQGDREKAY
NALIAECPLIPEYISDARVYTGNGELRYKIDQKSYHOTKWRGGMVLVGAACFVD
VFSGVHLATYASLALARSINSVLAGLDLDEKTADEFEGRRRREYGFYERLAFYOM
NVNESYFMIAKTYTNNEHTREESFVDVLVGVASGRTALAASTRABESAFRAGAVDM
AGSDGMVPLFSHVAVVQWMBGQVQKRAVLGSDAPAEVPEFGGLVSSPDGKML
PYRG"
complement (9077. .10273)
/note="P450-related oxidase"
/codon_start=1
/transl_table=1
/product="Stac"
/protein_id="AAM80533.1"
/db_xref="GI:21911410"
/translation="MSGDDRPPIITLRGSPADELRAAGELTVRLSGSADAEHTWL
ATGADVROYLDGHTRESTRRPRRNDIEGGKVFRRPVLGNIMODVDPPTLRL
LAPGFTTRKTRRNAPRYTEQVTERLDEMBEGSPADILFELPADVPPVYLCEILLGVPR
DDRMFLQTLCHRLIDASLSGRRAAEARSRYLVTVARERKDPGDLIGMVAEHG
DVTYDELRGCVQMMLAGDNNISGMIGVLALRNPEQIALRGVPAERAVDEL
IRYLTVYAATPRTAIEDSTVGQVIVAGETVLCSLPTANRDPALLPADRLDVTRA
VPHVAFGHVGHHCIGALALRELRLIAYTALMRPRPDRLADPDGATEPRISTPAYGIS
RLMYTW"
complement (10263. .11417)
/note="P450-related oxidase"
/codon_start=1
/transl_table=1
/product="Stac"
/protein_id="AAM80534.1"
/db_xref="GI:21911411"
/translation="WALPLPHRRRLDVPVEPHDLQNGEPLHEDYTPGMDGRKQMLV
TGHEVREIIEADEPERSFMSRVDDEARALLPGILOAYDPDHLRLRYTAPASAR
MERLRPRIDIEELDECDLSDGSPVDFVYVAMPIDALLACEFLDVRDRAALISM
IRBSERBRLPRORTSSGSGVNTYQKAAKRKLDPGGGMGLVYREAGAEVSDGLNP
LAEGLNLMAAEQAAQLAVALLVTHRPDMALLRHRBELVDGATEBLRHASIVLEAP

CDS

APRVLEDEVSVAGRDIRAGDVLTCSSMAVNRPOSEHPIITENPKHMAFGIHCIG
APLRLQIRVALPAVLRFRFSLRLVAPEDLRFKRGAPPAVELEPVM"
complement(11438.12613)
/note="P450-related oxidase"
/codon_start=1
/transl_table=11
/product="star"
/protein_id="AA080535.1"
/db_xref="GI:21911412"
/translation="MFPEINVRASQLHRDRPDPVPELHSMKEGGLTVIGTSDSTE
GRTAMLATGIDEVROVLSDEKFSARLITWPGFLOTQYDPPEHTRLRVVV
AFSHRMQKRPVRVIOVDSLDITESTGCVDPVHGMATPATCDFLIPDDO
ADLARILLASTPDSKRTAAGNKPMYKHOAOSRSGDDLFGVGGENDALIT
DAELTGVAFVWGAADQVARIILAGAMLMYQRAQRIILREKPTPEMLDTRIL
TTDEKTHPRVATQVIRIGNOLVKAGDVTCSLLANRPNVSADEPDTREKHEHIA
FPHGIHICIGRAMELMKVSI PALAHRFTLRILADPQREITLGPPEPVALLLDW"

Alignment Scores:

Pred. No.: 5.72e+03 Length: 69301
Score: 110.50 Matches: 73
Percent Similarity: 39.43% Conservative: 37
Best Local Similarity: 26.16% Mismatches: 96
Query Match: 7.56% Indels: 75
DB: 1 Gaps: 16

US-09-087-136-1 (1-275) x STU82965 (1-69301)

QY 2 SerGluLeuAspProLeuAlaGluPheLeuProGluAspGlyAspArgAsnAlaArg 21
DB 11956 GCCGAGCTGACGGGGGTGGCCGCTTCATGAGCGCGCTGCCGCCGACACAGGTGGCCGT 11897
QY 22 GluAsnAspProLeuLeuSerGlyGlyProLeuProLeuGluSerProSerArgGlyLeu 41
DB 11896 CTC-----CTGCGCGCGCGCGCGCTGATGTCGAACACCGCGCGAG----- 11852
QY 42 ThrSerLeuLeuSerTyrAspPro---ThrValProGluSerProAsp-----MetLys 58
DB 11851 TTCGCTCTGCTCAGGAGAAAGCCCGACCGTACCGAGTGGCTGAGAGACCATGGGG 11792
QY 59 PheAlaArgLysArgLeuGlyAsnLeuLeuThrTrpIleLysHisProSerGluIle 78
DB 11791 TAC-----CTACGACCGCA--CGAGAGACCCATCC----- 11763
QY 79 IleGlyValLeuProGluAspTyrThrArgAlaAspGluGluProGlyArgGluArg 98
DB 11762 -----GCGCGTGGGACCCAGACGTCGCGATCGGAAACGACTGTCAGAGCCGG 11712
QY 99 -----ProProGlyArg---ProArgLysMetPro----- 107
DB 11711 CGACACCGTCACTGCTCTGCTGCGCCGCAACCGCCGCAACTGCCGAGTGGCGAGGA 11652
QY 108 -----ArgHis-----GluSerSerThrSerLeuMetGluSerProArgLysThr 122
DB 11651 CGAGTTCCAGATCACAACGAGGAGAAAGCGAGCATCTGGCGTTCGGGACCGCATCCATCA 11592
QY 123 MetThrArgAspSerLysIleMetPheGluLeuArgGlyLysProPheGluMetIleAla 142
DB 11591 CTGCTCGGACAGGCCAT----- 11574
QY 143 GlyArgPheGluGluGluTyrSerLeuGlyArgAlaIleValIleGlyHisMetAsnAla 162
DB 11573 -----GGCCGAGCTGATGTTCAAGGTCATCGATTCCGGC 11541
QY 163 GluTyrGluProIle-----LysAlaGluArgThrAspTyrAlaProAsnLeuAla 179
DB 11540 GCTGGCGACCGCTTCCCAACCTCGGCTCGCATCCGACCGCGAGATCACCTTGGG 11481
QY 180 ValAspTyrLeuAlaCys---ArgGluIleHisArgMetProArgProAspLysSerIle 198
DB 11480 GCCGCCGCCGTTGCAATGGAAGCCCTGCTGCTCACTGGTATGCACCGGACGAGAGAGTA 11421
QY 199 ProGluLeuProIleValProSerArgIleAspGluPheAspAlaThrValAspProArg 218

DB 11420 CCAGTGGCACTTCCTTCGCGCAACCGGCGGACCGGCTCATCTCTG---CCGAG 11367
QY 219 TyrGluThrAspLeuLysAsnGluTyrIleArgHisTrrplysGluValLysGly--- 237
DB 11366 TTCAC---GACCTGCAAGACGAGGGCCGCTGCACAAATACGACACGAGCCGGGATG 11310
QY 238 -----TrrCysAlaIleGluArgArgArgThrAlaProHisAlaArg 251
DB 11309 GACGCGCGCTAAGACAGTGGCT--GGTCAACGGGTACACGAGGTTCGCACTCTCCGC 11254

RESULT 15

AB017282.24

WPCOMMENT

Sequence split into 33 fragments LOCUS AB017282 Accession AB017282

Fragment Name	Begin	End
AB017282_00	1	110000
AB017282_01	100001	210000
AB017282_02	200001	310000
AB017282_03	300001	410000
AB017282_04	400001	510000
AB017282_05	500001	610000
AB017282_06	600001	710000
AB017282_07	700001	810000
AB017282_08	800001	910000
AB017282_09	900001	1010000
AB017282_10	1000001	1110000
AB017282_11	1100001	1210000
AB017282_12	1200001	1310000
AB017282_13	1300001	1410000
AB017282_14	1400001	1510000
AB017282_15	1500001	1610000
AB017282_16	1600001	1710000
AB017282_17	1700001	1810000
AB017282_18	1800001	1910000
AB017282_19	1900001	2010000
AB017282_20	2000001	2110000
AB017282_21	2100001	2210000
AB017282_22	2200001	2310000
AB017282_23	2300001	2410000
AB017282_24	2400001	2510000
AB017282_25	2500001	2610000
AB017282_26	2600001	2710000
AB017282_27	2700001	2810000
AB017282_28	2800001	2910000
AB017282_29	2900001	3010000
AB017282_30	3000001	3110000
AB017282_31	3100001	3210000
AB017282_32	3200001	3304554

Continuation (25 of 33) of AB017282 from base 2400001 (AB017282 Methylcococcus capsulatus)

Alignment Scores:

Pred. No.: 9.77e+03 Length: 110000
Score: 110.50 Matches: 60
Percent Similarity: 39.83% Conservative: 34
Best Local Similarity: 25.42% Mismatches: 81
Query Match: 7.56% Indels: 64
DB: 1 Gaps: 12

US-09-087-136-1 (1-275) x AB017282_24 (1-110000)

QY 82 LeuProGluAspTyrThrArgAlaAsp---GluGluProGlyArgGluArgProPro 100
DB 42331 CTGCGACGGCGCTGCGCATCGAGTGTGAGAGAAGAATCGGCAAGCGCGCCACAA 42390
QY 101 GlyArgProArgLysMetProArgHisGluSerSerThrSerLeuMetGluSerProArg 120
DB 42391 GGTCAAGCCCGCCCAAGTTCCCAAGG--CTGCGCGGAATACG-----CCGCGGCCACAG 42443
QY 121 LysThrMetThrArgAspSerLysIleMetPheGluLeuArgGly-----LysPro 137
DB 42444 TCGACCTC-CAGCCCGCGTAATTCATTCGCGCTGCGGCTGCGGCGACTGGGAAAAATCCG 42502
QY 138 PheGluMetIleAlaGlyArgPheGluGluGluTyr-----SerLeuGlyArgAlaIleTrp 155

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2005, 06:21:50 ; Search time 84 Seconds

(without alignments)
1676.450 Million cell updates/sec

Title: US-09-087-136-1

Perfect score: 1462
Sequence: 1 MSEIDPLAEFLLPEDGDNRN.....INKIYGPESKTVQALGLI 275Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 03:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1462	100.0	275	2 Q23482	Q23482 caenorhabdi
2	111.5	7.6	337	2 Q8GYK6	Q8GYK6 arabidopsis
3	109	7.5	332	2 Q9MAU1	Q9MAU1 arabidopsis
4	104.5	7.1	717	2 Q83Z59	Q83Z59 streptomyce
5	102.5	7.0	798	2 Q9P6A7	Q9P6A7 neurospora
6	102.5	7.0	1054	2 Q9JM39	Q9JM39 mus musculu
7	101.5	6.9	1593	2 Q7SB29	Q7SB29 neurospora
8	100	6.8	301	2 Q87CR4	Q87CR4 xyliella fas
9	100	6.8	1766	2 Q7SE06	Q7SE06 abduya goss
10	98.5	6.7	1183	2 Q9RIV2	Q9RIV2 rattus norv
11	98	6.7	718	2 Q6GPD1	Q6GPD1 xenopus lae
12	98	6.7	1012	2 Q7YY92	Q7YY92 cryptospori
13	97.5	6.7	1257	2 Q9JLL7	Q9JLL7 mus musculu
14	97.5	6.7	1257	2 Q9WVR3	Q9WVR3 rattus norv
15	96.5	6.6	425	2 Q6K9A6	Q6K9A6 oryza sativ
16	96.5	6.6	678	1 ABPP_RIPCL	ABPP_RIPCL
17	96	6.6	567	2 Q7RTF7	Q7RTF7 homo sapien
18	96	6.6	1246	2 Q9P3E9	Q9P3E9 neurospora
19	95.5	6.5	393	2 Q7RJX5	Q7RJX5 plasmodium
20	95.5	6.5	448	2 Q6ZM26	Q6ZM26 streptomyce
21	95.5	6.5	531	1 GNT3_HUMAN	GNT3_HUMAN
22	95.5	6.5	533	1 Q6IC49	Q6IC49 homo sapien
23	95.5	6.5	1847	1 RIP1_RAT	RIP1_RAT
24	95	6.5	337	1 Y66L_SYNY3	Y66L_SYNY3
25	95	6.5	425	2 Q8UEN6	Q8UEN6 synecocyst
26	95	6.5	429	2 Q7CYR2	Q7CYR2 agrobacteri
27	95	6.5	429	2 Q9W2X8	Q9W2X8 dirosophila
28	94.5	6.5	379	2 Q9X780	Q9X780 mycobacteri
29	94.5	6.5	410	2 Q9CB10	Q9CB10 mycobacteri
30	94.5	6.5	479	2 Q08611	Q08611 mus musculu
31	94.5	6.5	1257	2 Q6P549	Q6P549 mus musculu

32	94	6.4	540	2 Q6GME0	Q6GME0 xenopus lae
33	93.5	6.4	192	2 Q8L4Z1	Q8L4Z1 arabidopsis
34	93.5	6.4	383	2 Q8NR26	Q8NR26 corynebacte
35	93.5	6.4	632	1 GIDA_STRP3	GIDA_STRP3
36	93	6.4	218	2 Q6RGN4	Q6RGN4 streptomyce
37	93	6.4	325	2 Q9W584	Q9W584 dirosophila
38	93	6.4	381	2 Q9VOD6	Q9VOD6 dirosophila
39	92.5	6.3	602	2 Q7MUS1	Q7MUS1 porphyromon
40	92.5	6.3	809	2 Q8RMQ1	Q8RMQ1 arabidopsis
41	92	6.3	296	2 Q67XS0	Q67XS0 arabidopsis
42	92	6.3	325	2 Q8SXL0	Q8SXL0 dirosophila
43	92	6.3	396	2 Q6D615	Q6D615 erwina car
44	92	6.3	520	2 Q6ZM33	Q6ZM33 homo sapien
45	92	6.3	589	2 Q9NP00	Q9NP00 homo sapien

ALIGNMENTS

RESULT 1

ID	Q23482	PRELIMINARY;	PRT;	275 AA.
AC	Q23482;			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-OCT-2001 (TREMBLrel. 18, Last sequence update)			
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)			
DE	Abnormal cell lineage protein 37.			
GN	Name=lin-37; ORFNames=ZK418.4;			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;			
OC	Rhabditidae; Peloderinae; Caenorhabditis.			
OX	NCBI_TaxID=6239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Bristol N2;			
RX	MEDLINE=99069613; PubMed=9851916;			
RG	WormBase Consortium;			
RT	"Genome sequence of the nematode C. elegans: a platform for			
RT	investigating biology. The C. elegans Sequencing Consortium.";			
RL	Science 282:2012-2018 (1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Bristol N2;			
RA	Fulton L.;			
RT	"The sequence of C. elegans cosmid ZK418.";			
RT	Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Bristol N2;			
RA	Waterston R.;			
RT	Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Bristol N2;			
RA	Wilson R.;			
RT	Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.			
RN	[5]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Bristol N2;			
RG	WormBase Consortium;			
RT	Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; U00047; AAA50688.2; -.			
DR	PIR; T27850; T27850.			
DR	IntAcc; Q23482; -.			
DR	WormBase; WBGene00003022; ZK418.4.			
DR	Mormep; ZK418.4; CB25690.			
DR	SEQUENCE 275 AA; 31601 MW; OPE64938F5A6107D CRC64;			
QY	Query Match	100.0%;	Score 1462;	DB 2; Length 275;
	Best Local Similarity	100.0%;	Pred. No. 6.1e-110;	
	Matches 275; Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
	1 MSEIDPLAEFLLPEDGDNRNQNPDPLISGGLPESPBRKLTSLSYPTVPESPDMKFA	60		

```

Db      1 MSEIDPLAEFLPEGDGNARNDPLISGRLPLESPSRKLTSLSTYPTVPESPMKFA
Qy      61 RRLGNLLTTIKHMSSEIIGVLPEDYTRADEBPGRQGRPKRPMRHSSTLSMSPR 120
Db      61 RRLGNLLTTIKHMSSEIIGVLPEDYTRADEBPGRQGRPKRPMRHSSTLSMSPR 120
Qy      121 KMTWDSKIMFELRGKPEMTAGREPEEYSIGRAMVKGHNNEPEPTKAORTDYPNLAV 180
Db      121 KMTWDSKIMFELRGKPEMTAGREPEEYSIGRAMVKGHNNEPEPTKAORTDYPNLAV 180
Qy      181 DYLACREIHRMPRDKSIPELPIVPSRIDEPDATTVDPRYETDLKNEYIRHMKOVKGCMA 240
Db      181 DYLACREIHRMPRDKSIPELPIVPSRIDEPDATTVDPRYETDLKNEYIRHMKOVKGCMA 240
Qy      241 HORRIAPHARSILINKIYOGESKTYEQALGI 275
Db      241 HORRIAPHARSILINKIYOGESKTYEQALGI 275

```

RESULT 2

```

Q8GYK6 ID O8GYK6 PRELIMINARY; PRT; 337 AA.
AC O8GYK6;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE Hypothetical protein At5g32840.
GN Name=At5g32840;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
RA Nakajima M., Enju A., Kamuya A., Narusaka M., Carninci P., Kawai J.,
RA Hayashizaki Y., Shinzaki K.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,
RA Onodera C.S., Quach H.L., Tang C., Toriumi M., Wong C., Wu H.C.,
RA Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
RA Ishida J., Jones T., Kamuya A., Kawai J., Kim C.J., Narusaka M.,
RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shim P.,
RA Southwick A., Tripp M.G., Wu T., Shinzaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK117550; BAC4221.1; -
DR EMBL; BT005099; AA050632.1; -
DR InterPro; IPR000425; MIP.
DR PROSITE; PS00221; MIP; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 337 AA; 37383 MW; 2D74924311C8C224 CRC64;

```

Query Match Best Local Similarity 7.6%; Score 111.5; DB 2; Length 337;
Matches 62; Conservative 40; Mismatches 90; Indels 71; Gaps 15;

```

Qy      18 RNARONDELISGGLPLSPSRKLTSLSTYPTVPESPMKFAKRLGNLLTTIKHPSR 77
Db      111 RPRVQNSVAD---PVGSPPRG-----GYTRPRG-----VYGHHGQ 145
Qy      78 IIGVLP--EDYTRADEBPGRQGRPPG-----RPRKMRHSSTLSMSP--RKT 122
Db      146 FVSNLDPMNQPMRAHPOQOSPOLGSGHMKGVHFLQPRATP---SPISILNSGHKA 202
Qy      123 MTRDSKIMFELRGKPEMTAGREPEEYSIGRAMVKGHNNEPEPTKAORTDYPNLAV 182
Db      203 RSRDPAVL-VRKRAKAITRG--ASLYSLCRSWLR--NGAHGKIKFOR-----IDM 248

```

```

Qy      183 LACREIHRMPR---DKSIPELP--IYPSRIDEPD---ATVDPRYETDLKNEYIRHMKQ 233
Db      249 MTC-----LPKLPVDKTETSLPDKLVEBALCEEDKEDEESVKHLSDELKRIHIDRAKK 303
Qy      234 VKKGWCAHQRRTPARHSIALI 256
Db      304 VRARLEERLKRIRAKARLALL 326

```

RESULT 3

```

Q9MAU1 ID Q9MAU1 PRELIMINARY; PRT; 332 AA.
AC Q9MAU1;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE F13M7.8 protein (Hypothetical protein).
GN Name=F13M7.8;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
[1]
RN SEQUENCE FROM N.A.
RA Vysotskaya V.S., Schwartz J.R., Toriumi M., Yu G., Kwan A., Liu S.,
RA Li J., Kremenevskaya I., Lueros J., Araujo R., Au M., Bredel V.,
RA Buehler E.E., Conway A., Dewar K., Feng J., Kim C., Kurtz D., Li Y.,
RA Palm C., Shinn P., Sun H., Davis R., Ecker J., Federspiel N.,
RA Theologis A.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RA Theologis;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RA MEDLINE=22088475; PubMed=12093376;
RA Haas B.U., Volkovsky N., Town C.D., Troukhan M., Alexandrov N.,
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT "Full-length messenger RNA sequences greatly improve genome
RT annotation."
RL Genome Biol. 3:RESEARCH0029-RESEARCH0029(2002).
[4]
RN SEQUENCE FROM N.A.
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC004809; AAF0444.1; -
DR EMBL; AY085490; AAM62716.1; -
DR PIR; G86182; G86182.
KW Hypothetical protein.
SQ SEQUENCE 332 AA; 36793 MW; 1B8B687141A070F4 CRC64;

```

Query Match Best Local Similarity 7.5%; Score 109; DB 2; Length 332;
Matches 69; Conservative 37; Mismatches 99; Indels 82; Gaps 16;

```

Qy      5 DP---LAEFLPEGDGNARQ-----NDPLISGGLPLESPSRKLTSLSTYPTVPESPD 56
Db      86 DESAVLYPFAFGRG-FSARVPRGVADPSVTAGNLGGYPRPSPFT-----YDPGYEQRQ 140
Qy      57 M-----KPARKRLGNLLTTIKHPSSEIIGVLPEDYTRADEBPGRQGRPPG----- 101
Db      141 MESLILQPTIRERNPQI-----RPLPRIGL-----GSPVIGPTIRASPOF 179
Qy      102 RPRKMRHSSTLSMSP--RKTWDSKIMFELRGKPEMTAGREPEEYSIGRAMVKG 158
Db      180 LQPRVAP---PPTSLDTSRNRKARSKDGLAV-VGRKRVITREGS--SLYSLSGRSLK- 233
Qy      159 HNNNEVEPTKAORTDYPNLAVYLACREIHRMPR-----DKSIPELPIVPSRID- 209
Db      234 --NGAHGKIKFORSGI-----MKPLKPLPVDLTETSTVDDPDDESSADEDK 278

```


InterPro: IPR010781, DUF1376.
DR Pfam: PF07120; DUF1376; 1.
KM Complete proteome; Hypothetical protein.
SQ SEQUENCE 301 AA; 34399 MW; F5B59399C8040E1D CRC64;

Query Match 6.8%; Score 100; DB 2; Length 301;
Best Local Similarity 25.2%; Pred. No. 6.5;
Matches 61; Conservative 19; Mismatches 80; Indels 82; Gaps 13;

QY 81 VLPRD---YTRADPEPRG-----QGRPPRPRKMPRHSESTSLMESP----- 119
DB 75 LLQEDGWNRKCDDEIARFARAAAQOENKGGKGRNSAHHDSDOTQOKPADFDLNP 134
QY 120 ----RKTWTRDSKIMFELRGKPE-----MIAGRFEEYSL-----GRAWKGMNNEY 164
DB 135 NETOSKANTNENETQIEPNQKPTRHQTPDPISSLSCEESLVFPAEDGNA---TSGTG--- 189
QY 165 EPIKORTDYAPNLAVDYIACREIHRMRPDKSIPELPIVPSRIDEPDAT---VDPRYET 221
DB 190 ---KPKSPHGSRLPDDW-----VPSVDVLATQOGVDGRYEA 225
QY 222 DLKNEYIRHWKQV-----KKGCAHQF---RTNPAHRSIALINKIYQGESKTVESQAL 272
DB 226 EKFRDY---WRSVAGAKGRKODWEATWMTNRRAEDKTS---SMKHGYQRHEYNRSRPMRL 280
QY 273 GL 274
DB 281 GL 282

RESULT 9

Q75E06 PRELIMINARY; PRT; 1766 AA.

AC Q75E06; 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE ABL133CP.
GN ORFNames=ABL133C.
OS Ashbya gossypii (Yeast) (Eremothecium gossypii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Eremothecium.
OX NCBI_TaxID=33169;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10895;
RA Voegelé S.E., Dietrich F.S., Brachat S., Lerch A., Gaffney T.,
RA Philippsen P.;
RL Submitted (SEP-2004) to the EMBL/Genbank/DBJ databases.
DR EMBL; AEO16815; AAS50638.1; -.
DR AGD; ABL133C; -.
SQ SEQUENCE 1766 AA; 179132 MW; BEC73A1CF69F2BA3 CRC64;

Query Match 6.8%; Score 100; DB 2; Length 1766;
Best Local Similarity 23.8%; Pred. No. 67;
Matches 56; Conservative 31; Mismatches 122; Indels 26; Gaps 11;

QY 3 EIDELAEFLP-EDGDRNARNDPLISGRLPLESPSRKLTSLSYDPTVESPDMPKPAR 61
DB 1123 ESKPATEYSTAPAGSTIILPGSPLPGTVPAGKSKRTATPRS--PAQVKNQDTKPSD 1180
QY 62 KRLGNLTITIGHNPSSEIGVLPEDYTRADEEPRGQRPGRPRKMPRHESSTSL---MES 118
DB 1181 KSTPATETSTP-TPEGSTPLPGS-PLLPHTPVASGKSESPASVPKKEDEVTEKKDDVRS 1238
QY 119 PRK---TWTRDSKIMFEL-RGKPFEMLAGRFEEYSLSGRAWKGMNNEY-EPIKORTD 173
DB 1239 PAKNAATVLSPLAAPVPKGPLVKT-----SVNQPSKESPGSEYTTPLATSPVD 1291
QY 174 YAPNLAVDYIACRE---IHRMRPDKSIPELPIVPSRIDEPDATVPRYETDLKN 225
DB 1292 VTHSLSGSAPVKKPVGVTAIPVASKSVNPVDKVEKPVNK---NLQPSKGTDTNN 1343

RESULT 10

Q9RLV2 PRELIMINARY; PRT; 1183 AA.

ID Q9RLV2 01-MAY-2000 (TREMBLrel. 13, Created)
AC Q9RLV2 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE SHIP-2.
GN Name=SHIP-2;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=99310603; PubMed=10381377; DOI=10.1006/dbrc.1999.0888;
RA Ishihara H., Sasaoka T., Hori H., Wada T., Hirai H., Haruta T.,
RA Kobayashi M.;
RT "Molecular cloning of rat SH2-containing inositol phosphatase 2
RT (SHIP2) and its role in the regulation of insulin signaling.";
RL Biochem. Biophys. Res. Commun. 260:265-272(1999).
DR EMBL; AB011439; BAA01818.1; -.
DR HSSP; O60880; 1D1Z.
DR GO; GO:0004437; P:inositol or phosphatidylinositol phosphatase. . .; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000300; IPPC.
DR InterPro; IPR000980; SH2.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00017; SH2; 1.
DR PRINTS; PR00401; SH2DOMAIN.
DR PRODOM; PD000093; SH2; 1.
DR SMART; SM00128; IPPC; 1.
DR SMART; SM00252; SH2; 1.
DR PROSITE; PS50001; SH2; 1.
SQ SEQUENCE 1183 AA; 131155 MW; 1B31EPD454B34D0E CRC64;

Query Match 6.7%; Score 98.5; DB 2; Length 1183;
Best Local Similarity 22.4%; Pred. No. 52;
Matches 54; Conservative 22; Mismatches 84; Indels 81; Gaps 8;

QY 7 LAEFLPEDGDRNARNDPLISGRLPLESPSRKLTSLSYDPTVESPDMPKPARKRG 66
DB 886 LYEWISIDKDTGAKSAPSVLRGNQEHRSGRKPTSTEASCP----- 928
QY 67 LTTIKHHPSEIIGVLPEDYTRADEEPRGQRPGRPRKMPRHESSTSLMES----- 118
DB 929 -LSKLFEEF-----EKPPTRPPAPRAVRESLNRKLSSEGTPEQEG 972
QY 119 ---PRKTWTRDSKIMFELRGKPFEMLAGRFEEYSLSGRAWKGMNNEYEPKORTDYA 175
DB 973 VAAPPPKNSFNPNAYVLEGVHQLP---LEPTSFARALIPPTKN-----KVALTVPA 1024
QY 176 PNLAVDYIACREIHRMR-----PKSIFELP-----IYPSRIDEPDAT 214
DB 1025 FQLG-----RRTRVVGSSSDSDSGTLPPPPPLPDSALFLPPNLDPLSMP 1076
QY 215 V 215
DB 1077 V 1077

RESULT 11

Q6GPD1 PRELIMINARY; PRT; 718 AA.

ID Q6GPD1 05-JUL-2004 (TREMBLrel. 27, Created)
AC Q6GPD1 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DE Mad1-A protein.
GN Name=mad1-A;

OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8335;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22386857; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feirngold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stempleton M., Soares M.B., Bonaldo M.F., Casavant P.L., Scheetz T.E.,
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carrinci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Molloy S.J.,
RA Bosak S.A., McGowan P.J., McKernan K.U., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A.C., Rodrigues S., Sanchez A.,
RA Whiting R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Blakesley R., Rodriguez A.C., Grimwood J., Schmutz L.Y., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Merrit M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research. The NIH Xenopus
RT initiative";
RL Dev. Dyn. 225:384-391(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Strausberg R.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL, BC073210; AAH73210.1; -
DR InterPro; IPR006672; MAD.
DR Pfam; PF05557; MAD; 1.
SO SEQUENCE 718 AA; 83248 MW; C82ECDBB851C7789 CRC64;

Query Match	6.7%;	Score 98;	DB 2;	length 718;
Best Local Similarity	26.4%;	Pred. No. 30;		
Matches	61;	Conservative	24;	Mismatches 82;
				Indels 64;
				Gaps 9
QY	46	SYD-----PTVPESPMKFARKLGNLLTKIHPSEIIGVLPEDYTRADEEGRQGRPPG	101	
Db	418	SYDSLTPT-EKSPQSRRLKAEIILQVQYHNAMETQSE----	ALFDGIGQK---	469
QY	102	RPRKMRHSSST-----SLMESPRKTTRDSKIMFELRGKPFEMIA--GRFEE----		148
		:	:	:
Db	470	-----KSELTLAEIKLVKSGQSSDQMTSPFNEMASMLRLKIELEBAERGRLEBKNTL	523	
QY	149	-----YSIGRAWKGNHNNEXPEPIAORTDYAPNIAVDYLACREIHMRRDPK	196	
Db	524	EMRLSELNLOGCYDSRFTVILSLNPASKAKQKQRTDVRHILQECDDKLEIVRLILEGA	583	
QY	197	SIPELPIVPSRIDEFDATVDPK-----YETDLKNEYITHMKQV	235	
Db	584	QIP-----DKLEATGSPQSSQSELAEIKKQVESAEIKXORLEEVQTK	625	
RESULT 12				
Q7YY92				
ID	Q7YY92	PRELIMINARY;	PRT,	1012 AA.
QC	Q7YY92;			

DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Putative arabinogalactan protein, possible.
GN ORFNames=IMB_373;
OS *Cryptosporidium parvum*.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC *Cryptosporidiidae*; *Cryptosporidium*.
OX NCBI_TaxID=5807;
RN [1]
RP SEQUENCE FROM N.A.
RA Bankier A.T., Spriggs H.F., Fartmann B., Konfortov B.A., Madera M.,
RA Vogel C., Teichmann S.A., Ivens A., Dear P.H.;
RT "Integrated mapping, chromosomal sequencing and sequence analysis of
RL *Cryptosporidium parvum*.";
RL Genome Res. 0:0-0(2003).
EMBL; BX558351; CAD98593.1;
SQ SEQUENCE 1012 AA; 111471 MW; BF162972BF743BE1 CRC64;

```

Query Match      6.7%; Score 98; DB 2; Length 1012;
Best Local Similarity 22.9%; Pred. No. 47;
Matches 58; Conservative 31; Mismatches 92; Indels 72; Gaps 12

QY :
   1 MSEIDPLAEFLIPEDGGDRNARNDPLISGG-PLP--LESPSRKTLSLYDPVPVESPDM 57
     ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5S2 MSTTNVEIITKPIHDVDYIRSDXYISKCVIPELPGSPSRGSGSPIS-DRIYIPSTDE 650
                                     ||| | | | | | | | | | | | | |
QY 58 KFAKKRLGNLTTHKHHPSEIIIGVLPEDYTRADEEPRQG---RPGRPKMRHESSTS 1144
     I FKP L R -----T S K D Y Q P F R K G S L H P P - T P E N L P D K A S G R G
Db 651 IFKPLR-----TSKDYPFRKGSLHP-TPENLPDKASGRG 6866
                                     ||| | | | | | | | | | | | | |
QY 115 LMESPRKMTDSDKMFELRGKPPFEMIAGRFEFEYSLGRA-----WVGKMNNY-- 164
     ::||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 687 FVK-----LQHKMKPRE---YETEFYSSGSVPSPOLTYFKRVSEREYNH 729
                                     ||| | | | | | | | | | | | | |
QY 165 -----EPKAKRTDYAPNLAVDYLAEREIHMPRPKSIPELPYV--PSRIDEF-- 211
     ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 730 DKTIKAKEBQKLEKKBQRMTGTGYKDLRGVELDGADSVSELFDISEBPSPLEPFE 789
     ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 212 -DATVPRYEIDL 223
     : : : : :
Db 790 PERIIPSKFETKL 802
     : : : : :

RESULT 13
Q9ULL7 PRELIMINARY; PRT: 1257 AA.
AC Q9ULL7 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE SH2-containing inositol 5-phosphatase 2.
GN Name=Imppl1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]_
RP SEQUENCE FROM N.A. MEDLINE=20079159; PubMed=10610720; DOI=10.1006/geno.1999.5995;
RA Schumann S., Carrio R., Behrends J., Pouillon V., Merino J.,
RA Clement S.;
RT "The mouse SHIP2 (Imppl1) gene: complementary DNA, genomic structure,
RT promoter analysis, and gene expression in the embryo and adult
RL mouse."; 62:260-271(1999).
RL Genomics
DR EMBL; AF162781; AAF28187.1; -.
DR HSSP; O60880; ID12.
MGD; MGI:1333787; Imppl1.
DR GO; GO:0004437; P:inositol or phosphatidylinositol phosphatas...; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000300; IPGC.
```

DR InterPro; IPR001660; SAM.
DR InterPro; IPR000980; SH2.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00536; SAM_1; 1.
DR Pfam; PF00017; SH2; 1.
DR PRINTS; PR00401; SH2DOMAIN.
DR ProDom; PD000093; SH2; 1.
DR SMART; SM00128; Ippc; 1.
DR SMART; SM00454; SAM; 1.
DR SMART; SM00252; SH2; 1.
DR PROSITE; PS50105; SAM_DOMAIN; 1.
DR PROSITE; PS50001; SH2; 1.
DR SEQUENCE 1257 AA; 138996 MW; E205F9DB827E50DD CRC64;
SQ
Query Match 6.7%; Score 97.5; DB 2; Length 1257;
Best Local Similarity 22.4%; Pred. No. 68;
Matches 54; Conservative 23; Mismatches 83; Indels 81; Gaps 8;
QY 7 LAEFLPEDDGRNARNDPLISGGPLPLESPSRKLTSLSYDPTVPESPDMKFAKRLGN 66
DB 886 LYEWISIDKDDTGAKSKVPSVRSQEHRSGRKPAETISCP----- 928
QY 67 LTTTIKHPSEIIIGVLPEDYTRADDEEPCRGPRGPRKMRH-----SSTSL 115
DB 929 -LSKLFEEPP-----EKPPPTGRPPAPRAVREBPLNRLKSEGTSEGEV 972
QY 116 MESRKMTTRDSKIMFELRGKPFEMIAGRFEEBYSLGRAWKGMNNEYEPIKAQRTDYA 175
DB 973 VAAPPKNSEFNPNPYVYLVGVPHQLP---LEPPLAAAPLPAPATKN-----KVAITVPA 1024
QY 176 PNLAVDYLACREIHRMPR-----PKSIPELP---IVPSRIDEFDT 214
DB 1025 PQLG-----RRTRRVGEGSSSDSDSGTLEPPDPPLPDSALFLPNLDPLSMR 1076
QY 215 V 215
DB 1077 V 1077
RESULT 14
Q9WVR3 PRELIMINARY; PRT; 1257 AA.
ID Q9WVR3
AC Q9WVR3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE SH2-containing inositol phosphatase 2 (SHIP2).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20115790; PubMed=10648902; DOI=10.1016/S0169-328X(99)00311-3;
RA Kudo M., Saito S., Owada Y., Suzuki H., Kondo H.;
RT "Localization of mRNA for SHIP2, SH2 domain-containing inositol
RT polyphosphate 5-phosphatase, in the brain of developing and mature
RT rats";
RL Mol. Brain Res. 75:172-177(2000).
DR EMBL; AB025794; BAA82308.1; -.
DR HSPR; O60880; I012.
DR GO; GO:0004347; P:inositol or phosphatidylinositol phosphatase. . .; IEA.
DR GO; GO:0007242; P:intracellular signalling cascade; IEA.
DR InterPro; IPR005135; Exo_endo_phos.
DR InterPro; IPR000300; Ippc.
DR InterPro; IPR001660; SAM.
DR InterPro; IPR000980; SH2.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR Pfam; PF00536; SAM_1; 1.
DR Pfam; PF00017; SH2; 1.
DR PRINTS; PR00401; SH2DOMAIN.
DR ProDom; PD000093; SH2; 1.

DR SMART; SM00128; Ippc; 1.
DR SMART; SM00454; SAM; 1.
DR SMART; SM00252; SH2; 1.
DR PROSITE; PS50105; SAM_DOMAIN; 1.
DR PROSITE; PS50001; SH2; 1.
DR SEQUENCE 1257 AA; 139141 MW; 3A994C8E52940083 CRC64;
SQ
Query Match 6.7%; Score 97.5; DB 2; Length 1257;
Best Local Similarity 22.4%; Pred. No. 68;
Matches 54; Conservative 22; Mismatches 84; Indels 81; Gaps 8;
QY 7 LAEFLPEDDGRNARNDPLISGGPLPLESPSRKLTSLSYDPTVPESPDMKFAKRLGN 66
DB 886 LYEWISIDKDDTGAKSKAPVLRSGEHRSGRKPETISACP----- 928
QY 67 LTTTIKHPSEIIIGVLPEDYTRADDEEPCRGPRGPRKMRHSSSTLMES----- 118
DB 929 -LSKLFEEPP-----EKPPPTGRPPAPRAVREBPLNRLKSEGTSEGEV 972
QY 119 ---PRKMTTRDSKIMFELRGKPFEMIAGRFEEBYSLGRAWKGMNNEYEPIKAQRTDYA 175
DB 973 VAAPPKNSEFNPNPYVYLVGVPHQLP---LEPTSFAPAPLPPTKN-----KVAITVPA 1024
QY 176 PNLAVDYLACREIHRMPR-----PKSIPELP---IVPSRIDEFDT 214
DB 1025 PQLG-----RRTRRVGEGSSSDSDSGTLEPPDPPLPDSALFLPNLDPLSMR 1076
QY 215 V 215
DB 1077 V 1077
RESULT 15
Q6K9A6 PRELIMINARY; PRT; 425 AA.
ID Q6K9A6
AC Q6K9A6;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein O01520_C09.22 (Hypothetical protein
DE O0112_G06.10).
GN Name=O01520_C09.22; Synonyms=O0112_G06.10;
OS Oryza sativa (Japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Erihartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AP004064; BADI9244.1; -.
DR EMBL; AP003996; BADI9129.1; -.
KW Hypothetical protein.
SQ SEQUENCE 425 AA; 46800 MW; 877A3DFA42ASBD706 CRC64;
QY 15 DGDNRARNDPLISGGPLPLESPSRKLTSLSYDPTVPESPDMKFAKRLGNLTIT-- 71
DB 201 DGEARQRTSPPKLAIYVLVH-----LDQVLDNRPPTPEPTPMNLSQASGIISPSMT 255
QY 72 ---KHPSEII-----GVLPEDYTRADDEEP-----RGRRPGRPRKMRHSS 112
DB 256 EEPKHN--DFIMSLGVKDGTLPPQ--PPAFPSPGGRDRSRSRSPAPERRRPLGMDV 312
QY 113 TSLMES--PRKMTTRDSKIMFELRGKPFEMIAGRFEEBYSLGRAWKGMNNEYEPIKAQ 170
DB 313 PPARSSAFPRRSSRSSE-----NRGSGYGGHRAACSPVRAA 351
QY 171 -----RTDYAPNLAVDYLACREIHRMPRDXSIPELP---VPSRIDEFDTVDREYE 220

Db 352 RSRSPRGSSPPPSLSVPAVMKDGFGQJGRPPSTSPQDPIQTQATPLQANQFVMEADPMLJ 411

Search completed: April 9, 2005, 07:56:19
Job time : 87 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2005, 07:35:05 ; Search time 24 Seconds
(without alignments)
1102.484 Million cell updates/sec

Title: US-09-087-136-1
Perfect score: 1462
Sequence: 1 MSEIDPLAEFLLPEDGDRNA.....INKTYGPESKTVQALGLI 275

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1262	86.3	237	2 T27850	hypothetical prote
2	109	7.5	332	2 G86182	hypothetical prote
3	96	6.6	1246	2 T51085	related to protein
4	95.5	6.5	531	2 JN0586	beta-1,4-mannosyl-
5	95	6.5	337	1 S76634	hypothetical prote
6	95	6.5	425	2 AG2787	long-chain fatty a
7	95	6.5	429	2 A97567	hypothetical prote
8	94.5	6.5	410	2 D87151	probable integral
9	92	6.3	606	2 T51880	hypothetical prote
10	91	6.2	309	2 T02613	hypothetical prote
11	91	6.2	346	2 T01123	hypothetical prote
12	91	6.2	651	2 T40459	hypothetical lysin
13	90.5	6.2	1274	2 T16251	hypothetical prote
14	90	6.2	808	2 A89825	conserved hypotet
15	90	6.2	221	2 T04092	phospholipase D (E
16	88.5	6.1	267	2 A46122	homeoic protein H
17	88.5	6.1	535	2 T45831	SP16-like protein
18	88.5	6.1	814	2 T26702	hypothetical prote
19	88.5	6.1	1704	2 A59188	ATP-binding caset
20	88.5	6.1	1704	2 S71363	probable ATP-bind
21	88.5	6.1	3216	2 C90538	hypothetical prote
22	88	6.0	455	2 B86427	hypothetical prote
23	88	6.0	839	2 T50590	class I INCEMP pro
24	88	6.0	1417	2 T21012	hypothetical prote
25	88	6.0	1592	2 T16055	hypothetical prote
26	87.5	6.0	301	2 C82667	conserved hypotet
27	87.5	6.0	536	2 A43415	beta-1,4-mannosyl-
28	87.5	6.0	1994	2 D86452	protein F6N18.13 f
29	87	6.0	446	2 T34782	probable signal pe

30	87	6.0	756	2 T27642	hypothetical prote
31	87	6.0	844	2 S61112	regulatory protein
32	87	6.0	1047	2 T16203	hypothetical prote
33	87	6.0	4957	2 T03455	ALK protein - huma
34	87	6.0	5262	2 T03454	ALK protein - huma
35	86.5	5.9	640	2 A43273	heregulin, splice
36	86.5	5.9	645	2 B43273	hypothetical prote
37	86.5	5.9	739	2 T49456	hypothetical prote
38	86.5	5.9	805	2 T24431	hypothetical prote
39	86.5	5.9	1109	2 A40801	phosphoprotein pho
40	86.5	5.9	2022	2 T48818	glucan 1,4-alpha-g
41	86	5.9	603	2 A84584	hypothetical prote
42	86	5.9	610	2 A49082	calcium-dependent
43	86	5.9	1140	2 A80180	probable membrane
44	85.5	5.8	605	2 S18648	protein kinase wis
45	85.5	5.8	2938	2 T30249	cell proliferation

ALIGNMENTS

RESULT 1
T27850
hypothetical protein ZK418.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T27850
R:Pulton, L.
submitted to the EMBL Data Library, April 1994
A:Description: The sequence of C. elegans cosmid ZK418.
A:Reference number: Z20430
A:Accession: T27850
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-237 <FOLD>
A:Cross-references: UNIPROT:Q23482; EMBL:U00047; PIDN:AAA50688.1; CESP:ZK418.4
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:ZK418.4
A:Introns: 42/2; 87/2; 180/3
C:Superfamily: Caenorhabditis elegans hypothetical protein ZK418.4

Query Match 86.3%; Score 1262; DB 2; Length 237;
Best Local Similarity 100.0%; Pred. No. 1.8e-97;
Matches 237; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MSEIDPLAEFLLPEDGDRNARQNDPLISGGPLPESPSRKLTSLSYDPTVPSPPDMKFA	60
Db	1	MSEIDPLAEFLLPEDGDRNARQNDPLISGGPLPESPSRKLTSLSYDPTVPSPPDMKFA	60
Qy	61	RKRIGNLITTIKHPSEIIGVLPEDYTRADEEPORQRPGRPKMRHSESTLSMSPR	120
Db	61	RKRIGNLITTIKHPSEIIGVLPEDYTRADEEPORQRPGRPKMRHSESTLSMSPR	120
Qy	121	KTMTRDSKIMPELKGPFEMTAGFESEYSIGRAVNGHMNNEYEP1KAORTDYAPNLAIV	180
Db	121	KTMTRDSKIMPELKGPFEMTAGFESEYSIGRAVNGHMNNEYEP1KAORTDYAPNLAIV	180
Qy	181	DYLACREIHRMPRDKSIPELPIVPSRIDEDATVDRPYETDLNEXYIRHKQYKKG	237
Db	181	DYLACREIHRMPRDKSIPELPIVPSRIDEDATVDRPYETDLNEXYIRHKQYKKG	237

RESULT 2
G86182
hypothetical protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: G86182
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federici, N.A.; Kaul, S.; White, O.; Alonso, Chiu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Hultzar, L.
Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
A:Reference number: A66141; MUID:21016719; PMID:11130712
A:Accession: G66182
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-332 <STO>
A:Cross-references: UNIPROT:Q9MAU1; GB:AE005172; NID:q7211973; PIDN:AAF40444.1; GSPDB:GN
C:Genetics:
A:Map position: 1

Query Match 7.5%; Score 109; DB 2; Length 332;
Best Local Similarity 24.0%; Pred. No. 0.16;
Matches 69; Conservative 37; Mismatches 99; Indels 82; Gaps 16;
QY 5 DP--LAEFLPEDGDHNAQ-----NDPLISGGPPLBSPKRLTSLSYDPTVESPD 56
DB 86 DPAVAVLPALPGRG-PSARPVAGFVADPSVTAGNLSGVPRPSFT---YDFGPEYQRQ 140
QY 57 M-----KFAKRLGNLTITKHPSEITGVLPEDYTRADEBPGRQGRPRG----- 101
DB 141 MESLIQGFIREKRPDI-----RPLPLGLP-----GSPVGLGPIRASQPF 179
QY 102 -RRKMRHSESTSLMESP--RKTMTSDSKIMELAGKPEMIAAGFESEYSLGRAMVKG 158
DB 180 LQRRVAP--PPSILDTSNKRSKSDGALAV-VGKRVKRTIEGS-SSLYSLGRSMWK- 233
QY 159 HNNNEYPKIAQRTDYAPMLAVYLACREIHRMPR-----DKSIPFLPIVPSRID-- 209
DB 234 --NGAHVGIGIQPQSGI-----MKPLPKPLPVDTLTENISVPPDPDEBSADEK 278
QY 210 EPDATVDPREYTDLKNEXIRHMKQVKKCAQRRRTAFAASIALI 256
DB 279 EDEBAVKQLSEKDLKRIERAKVQAQLREERSRIRRYKERTILI 325

RESULT 3
T51085
related to protein kinase PAK1 [imported] - Neurospora crassa
N:Alternate names: protein B2A19.190
C:Species: Neurospora crassa
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 09-Jul-2004
R:Schulte, U.; Aign, V.; Hohelsel, J.; Brandt, P.; Fattmann, B.; Holland, R.; Nyakatura,
A:Reference number: 225286
A:Accession: T51085
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1246 <SCH>
A:Cross-references: UNIPROT:Q9P3E9; EMBL:AL390092; GSPDB:GN00116; NCSP:B2A19.190
A:Experimental source: BAC clone B2A19; strain OR74A
C:Genetics:
A:Gene: NCSP:B2A19.190
A:Map position: 6
A:Introns: 95/2

Query Match 6.6%; Score 96; DB 2; Length 1246;
Best Local Similarity 24.4%; Pred. No. 11;
Matches 63; Conservative 31; Mismatches 90; Indels 74; Gaps 17;
QY 48 DPTVP-ESPDMKFARKRLGNLTITKHPSEITGVLPEDYTRADEBPGR--QGRPPORP- 103
DB 2 EPQOPTQOPSMSSSSQ-----TPTQHPRPLHPLF---PMASQIPHRANSTPVSSG 52
QY 104 ---RKMPRHSESTSLMSPKTM---TRDSKIMFELAG-KPEMIAAGFESEYSLGRAMV 156
DB 53 LFSISIRSNMNSLGGPSESENTTAPATLHSPYLHPQLQSHKVRTHKANVEHDYTGKAI 112

QY 157 KGHNNXEYPIK-----AQRDYAPNLAVDL-----ACREIHRMPRPDKS 197
DB 113 -----NQVEIIEELRGHGKVKLAKNTQTGDVNAIKITPFPSKKRRLGKVTAMSTODKS 167
QY 198 IELPDIY-----PSRIDEFATVPDPRYETDKNEX--IRH-----WKQYKGM-----CA 240
DB 168 KREIATILKIRHPNVVALLEIIDDP-----ELKITYVLEHVELEBVVNR--KGLPLHICA 221
QY 241 HORRR-----TAPHAR 251
DB 222 YERRQEREQLGALPDAR 239

RESULT 4
JN0586
beta-1,4-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase (EC 2.4.1.144) - h
N:Alternate names: beta-D-mannoside beta1,4-N-acetylglucosaminyltransferase III; N-acety
C:Species: Homo sapiens (man)
C:Date: 24-Feb-1994 #sequence_revision 24-Feb-1994 #text_change 09-Jul-2004
R:Ikura, Y.; Nishikawa, A.; Tohma, T.; Soejima, H.; Nishikawa, N.; Taniguchi, N.
J. Biochem. 113, 692-698, 1993
A:Title: cDNA cloning, expression, and chromosomal localization of human N-acetylglucosa
A:Reference number: JN0586; MUID:93380894; PMID:8370666
A:Accession: JN0586
A:Molecule type: mRNA
A:Residues: 1-531 <IHA>
A:Cross-references: UNIPROT:Q09327; GB:D13789; NID:g398137; PIDN:BA02937.1; PID:g398138
A:Experimental source: fetal liver
C:Comment: This enzyme catalyzes the addition of N-acetylglucosamine in beta-1,4 linkage
ue.
C:Genetics:
A:Map position: 22q13.1
A:Keywords: glycoprotein; glycosyltransferase; hexosyltransferase; transmembrane protein
F:6-21/Domain: transmembrane #status predicted <TM>
F:169-180/Region: epidermal growth factor-like
F:139,239,257,395/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 6.5%; Score 95.5; DB 2; Length 531;
Best Local Similarity 22.6%; Pred. No. 4;
Matches 67; Conservative 27; Mismatches 90; Indels 113; Gaps 17;
QY 6 PLAEFLPEDGDRNANQNPILISGPT--PLSPSRKLTSLSYPTYPESPDMAFKAR 63
DB 53 PVPFQASPEPGGDLIRT-PLYSHSPLQLP-PPSKAAEHLRVLDVLPEDTTEVFVXTK 110
QY 64 LGNLTITKHPSEIIGVLPEDYTRADEBPGRPPGRPRKMPRHSESTSLMESPPKTM 123
DB 111 AG-----GVCFKPGTWMLEP-----PPGRPEKPPGANGSSARPPRYLL 151
QY 124 TRDSKIMFELRGRPREMIAAGFESEYSLGRAMV-----GHNNXEYPIKAQRTDYAPNL 178
DB 152 SAR-----ERTGR-----GARRKVEECVLPWG-----HGSC 180
QY 179 AVDYLAEREIHRMPRPDKSIPFLPIVPSRI-----DEPD----- 212
DB 181 GVPYIV--QYSNLPRKEHLVPR--EVRPRVAINAVNHEFLDVRFEHLGDVDAFYVC 236
QY 213 -----ATVDP--YETDLK--EYIRHMKOV-----KKGMCAHORRR 246
DB 237 ESNFTAYGEPRLPKREMLTNGTFEYIRH-KVLVYFLDHPFGRGQGDVADYLT 292

RESULT 5
S76634
hypothetical protein - *Synechocystis* sp. (strain PCC 6803)
C:Species: *Synechocystis* sp.
A:Variety: PCC 6803
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: S76634
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,
DNA Res. 3, 109-136, 1996

A:Reference number: Z14204
A:Accession: T00784
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-346 <ROM>
A:Cross-references: EMBL:AC003974; NID:g2914688; PID:g2914690
A:Experimental source: cultivar Columbia
R:Lin. X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffet, K.S.; Cronin, L.A.; Shen, M.; Vanden, S.E.; Umayam, U.; Tallon, L.;
Euse, D.; Nierman, W.C.; White, O.; Eisen, U.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: B84738
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-346 <STO>
A:Cross-references: GB:AE002093; NID:g2702278; PID:AB91981.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g32840; T21L14.22; F24L7.2
A:Map position: 2
A:introns: 185/3; 196/3; 214/2; 227/3; 253/3; 289/3; 316/2

Query Match 6.2%; Score 91; DB 2; Length 346;
Best Local Similarity 22.5%; Pred. No. 5.4;
Matches 61; Conservative 40; Mismatches 92; Indels 78; Gaps 16;

Qy 18 RNARQNDPLISGGLPSPSRKLTSLISYDTPVPSFPMKFAKRLGNLTTIKHPE 77
Db 111 RPYVNSNSVAD---PVGSPSPG---GYTPRGV-----VYGYHGG 145
Qy 78 IIGULP--EDYTRADEEGRCQ-----GRPRGRPKM-PR-----HESSTSLM 116
Db 146 FVSNLDPNNQPMRAHPNQQSPOLGSHMKGVPHFLQPRLVSENVYVATPSPTSL 205
Qy 117 ESP--RKTMTDSKIMFELRGKPFEMIGRFEESYSLGRAVWKGHMNNEPIKAQRTDY 174
Db 206 DNSGHKKARSDDALVL-VKRRKVRITEG--ASLYSLCRSMUR--NKAHGKIK----- 253
Qy 175 APNLAVDYLAEREIHRMRP--DKSIPELP-IVPSRIDEPD---ATVDPRYETDLKN 225
Db 254 ----RIDMWTG----LPKPLPVDKETSTSLPKDLVEAICEDKEDSESVGHLSDDLK 304
Qy 226 EYIRHMKYQKGCACGRRRTAPARSAIALI 256
Db 305 RHIDRAKKVRARLRERELKRIARYKARLAL 335

RESULT 12
T40459
hypothetical lysine rich protein - fission yeast (*Schizosaccharomyces pombe*)
C:Species: *Schizosaccharomyces pombe*
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T40459
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Baker, S.; Mungall, K.
submitted to the EMBL Data Library, November 1998
A:Reference number: Z21931
A:Accession: T40459
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-651 <LVN>
A:Cross-references: UNIPROT:O94356; EMBL:AJ034382; PID:CAA22282.1; GSPDB:GN00067; SPDB:
A:Experimental source: strain 972h-; cosmid c428
C:Genetics:
A:Gene: SPDB:SPBC428.07
A:Map position: 2
A:introns: 123/2; 215/3

Query Match 6.2%; Score 91; DB 2; Length 651;
Best Local Similarity 21.2%; Pred. No. 12;
Matches 52; Conservative 36; Mismatches 105; Indels 52; Gaps 10;

Qy 30 GPLPSPSRKLTSLISYD--TYVESPDMKFAKRLGNLTTIKHNSIIGVLPEVYT 87

Db 342 GKTPQNDTASPIDNETADPVDITVEAGSVENENETNGITTEHFA-----T 391
Qy 88 RADEPRGQGRPPGRPKMRPHESSTLMSFKTMTDRSKIMFELGKPFEMTAGRPE 147
Db 392 TEVAPASSEEKPPATGPA-----ESTSTQNVQASTQNDN-----GTPIGQIADALIE 439
Qy 148 EYSLGRAMVGHMNNNEPIKAQRTDYAPNLAVDYLAEREIHRMRPDKSIPELPIVPSR 207
Db 440 D-----VKGTAOS-VEQAFIETDIA-----LPDIALPDV-ATPTA 473
Qy 208 IDEPDATVDPRYETDLKNEYIRHMKYQKGCACGRRRTAPARSAIAL-INKIYQPGESK 266
Db 474 DDQDPSTAAATNEESVANE-DRTSKAGKGGHRRHKKKGGKMKVFGPKLHKYAKPTKSV 532
Qy 267 TVEQA 271
Db 533 KTLQA 537

RESULT 13
T16251
hypothetical protein F35A5.1 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T16251
R:Leimbach, D.
submitted to the EMBL Data Library, January 1996
A:Description: The sequence of C. elegans cosmid F35A5.
A:Reference number: Z18485
A:Accession: T16251
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1274 <LEI>
A:Cross-references: UNIPROT:Q20007; EMBL:U46675; NID:g1166613; PID:g1166621; PID:AB526
A:Experimental source: strain Bristol N2; clone F35A5
C:Genetics:
A:Gene: CESP:F35A5.1
A:Map position: X
A:introns: 1272/2

Query Match 6.2%; Score 90.5; DB 2; Length 1274;
Best Local Similarity 20.2%; Pred. No. 32;
Matches 58; Conservative 34; Mismatches 92; Indels 103; Gaps 13;

Qy 5 DPLAEFLP-EDGDRNARQNDPLISGGLPSPSRKLTSLISYDTPV----- 53
Db 668 NPVKMKRPWEDDDAPAK-----FVSLPBEKKTPLAKKAPTKPDSEAAADPVSG 718
Qy 54 --SPDMFAKRLGNLTTIKH--PSEIIGVLPEDYTA-----DEPRGQGRPPGR 102
Db 719 PSSKDPPLAKK-----APVPRDPSPMKAVPIPAKTEVPAPVVKPEVVASRPS- 771
Qy 103 PRKMPHRESSSTLMSFKTMTDRSKIMFELGKPFEMIGRFEESYSLGRAVWKGHMN 162
Db 772 PKKAKAPNSVFPPTVKKPVKMKRPWEDDDAPAPVNVPEBKKT----- 819
Qy 163 EYEPDKAQRTDYAPNLAVDYLAEREIHRMRPDKSIPELPIV----- 204
Db 820 ---PVLAKTLPVKP-----RDSPKKAVPAKPESTKTADAPVSVKPEPVSKEP 863
Qy 205 ----PSRIDEPDATVDPRYETDLKNEYIRHMK-----QYKK 236
Db 864 KEPSPKAEPSNPVPP--TPVKNP-VKWKRPWEDDEPTTEVKK 906

RESULT 14
A89825
conserved hypothetical protein SA0525 [imported] - *Staphylococcus aureus* (strain N315)
C:Species: *Staphylococcus aureus*
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: A89825
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2005, 07:42:10 ; Search time 26 seconds
(without alignments)
789.557 Million cell updates/sec

Title: US-09-087-136-1
Perfect score: 1462
Sequence: 1 MSEIDPLAEFLPEDGDRNA.....INKIYOGESKTVQALGLI 275

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/PCTUS.COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100	6.8	748	4	US-09-252-991A-18427
2	95.5	6.5	531	3	US-08-975-114A-4
3	95.5	6.5	531	3	US-08-849-281A-4
4	95	6.5	530	4	US-09-252-991A-28311
5	94.5	6.5	1037	4	US-09-252-991A-25361
6	91.5	6.3	2220	4	US-09-335-011-1
7	90.5	6.2	580	4	US-09-248-796A-20111
8	90	6.2	1780	1	US-08-769-309A-5
9	90	6.2	1780	3	US-08-994-570-5
10	90	6.2	1781	4	US-09-961-403-13
11	89.5	6.1	320	4	US-09-270-767-34209
12	89.5	6.1	320	4	US-09-270-767-49426
13	89.5	6.1	773	4	US-09-270-767-31141
14	89.5	6.1	773	4	US-09-270-767-48358
15	89.5	6.1	808	4	US-09-270-767-36557
16	89.5	6.1	808	4	US-09-270-767-51774
17	89.5	6.1	847	4	US-09-976-5594-401
18	89.5	6.1	853	4	US-09-949-016-11423
19	89	6.1	1467	4	US-09-252-991A-26888
20	88.5	6.1	1684	3	US-08-665-259-25
21	88.5	6.1	1684	3	US-08-762-500-25
22	88.5	6.1	1704	3	US-08-762-500-75
23	88.5	6.1	1766	4	US-09-949-016-10796
24	87.5	6.0	536	2	US-08-975-114A-3
25	87.5	6.0	536	2	US-08-849-281A-3
26	87.5	6.0	1704	4	US-09-032-438C-120
27	87	6.0	1257	2	US-08-750-152A-2

28	86.5	5.9	178	4	US-09-252-991A-28980	Sequence 28980, A
29	86.5	5.9	427	4	US-08-467-602-189	Sequence 189, App
30	86.5	5.9	427	4	US-08-411-295F-313	Sequence 313, App
31	86.5	5.9	430	4	US-08-467-602-187	Sequence 187, App
32	86.5	5.9	430	4	US-08-411-295F-311	Sequence 311, App
33	86.5	5.9	436	4	US-08-467-602-201	Sequence 201, App
34	86.5	5.9	436	4	US-08-411-295F-325	Sequence 325, App
35	86.5	5.9	439	4	US-08-467-602-198	Sequence 198, App
36	86.5	5.9	439	4	US-08-411-295F-322	Sequence 322, App
37	86.5	5.9	450	4	US-08-467-602-195	Sequence 195, App
38	86.5	5.9	450	4	US-08-411-295F-319	Sequence 319, App
39	86.5	5.9	459	4	US-08-467-602-204	Sequence 204, App
40	86.5	5.9	459	4	US-08-411-295F-328	Sequence 328, App
41	86.5	5.9	569	4	US-08-467-602-237	Sequence 237, App
42	86.5	5.9	569	4	US-08-411-295F-163	Sequence 163, App
43	86.5	5.9	572	4	US-08-467-602-235	Sequence 235, App
44	86.5	5.9	572	4	US-08-411-295F-161	Sequence 161, App
45	86.5	5.9	578	4	US-08-467-602-249	Sequence 249, App

ALIGNMENTS

```
RESULT 1
US-09-252-991A-18427
; Sequence 18427, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18427
; LENGTH: 748
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18427

Query Match      6.8%, Score 100, DB 4, Length 748,
Best Local Similarity 22.2%, Pred. No. 0.086,
Matches 54; Conservative 24; Mismatches 81; Indels 84; Gaps 13;

QY      74 HPSFLIGVLPEDYTRADEPGRGRRPRPKMKRHSSTSLMSPRKTMTRDSKIMREL 133
      139 HPG-----HOQHRGSSPTQOPRGRHRHPQPHQ-----LHRHPR-----W 178
      134 RGFEMFIAGFEESEYSGRAMVKGHMNEYEP1-----KAQRTYAPNL 178
      179 RYRL-----GARLRPRRGDTVRRH-----APVHRHGSLPQRSVGRHRRPRRGL 229
      179 AV-----DYLAQREIHRMPRDKSIFPL-----PIYV-----SRIDEFATYDP 217
      230 GAVRRRRRRRDQGAEEAFRRRDPQPPARLRLRPPVGRRLADROPOLFAQDEP 289
      218 RYEDLNKNEYIRHMKQYKKGCACOR-----BRTAPHASIALINKIYOGESKTVQ 270
      290 AAEPRLD---RPWRLAOPG---HQRGALLAGQRRSGIHRP-----RLMRPGADERLRH 337
      271 ALG 273
      338 PAG 340

RESULT 2
US-08-975-114A-4
; Sequence 4, Application US/08975114A
```

/ Patent No. 5876714
/ GENERAL INFORMATION:
/ APPLICANT: Atsushi NISHIKAWA et al.
/ TITLE OF INVENTION: HUMAN GLYCOSYLTRANSFERASE GENE,
/ NUMBER OF INVENTION: COMPOUNDS AND METHOD FOR INHIBITING CANCEROUS METASTASIS
/ NUMBER OF SEQUENCES: 5
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
/ STREET: 2033 K Street, N.W., Suite 800
/ CITY: Washington
/ STATE: D.C.
/ COUNTRY: U.S.A.
/ ZIP: 20006
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: MS-DOS
/ SOFTWARE: Wordperfect 5.1
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/975,114A
/ FILING DATE: No. 5876714ember 20, 1997
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/107,173
/ FILING DATE: August 17, 1993
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Lee Cheng
/ REGISTRATION NUMBER: 40,949
/ REFERENCE/DOCKET NUMBER: 1-F3439DIV
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 202-721-8200
/ TELEFAX: 202-721-8250
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 4:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 531 amino acid residues
/ TYPE: amino acid
/ STRANDEDNESS:
/ TOPOLOGY: linear
/ MOLECULE TYPE:
US-08-975-114A-4

Query Match 6.5%; Score 95.5; DB 2; Length 531;
Best Local Similarity 22.6%; Pred. No. 0.16;
Matches 67; Conservative 27; Mismatches 90; Indels 113; Gaps 17;
QY 6 PLAFLPLPDGDRNARNONDLISGGPL--PLESPSKLTSLSYDPTVPESDMKFAKR 63
DB 53 PVTPOASPPGPGDLRLT-PLYSHSPLQLPL-PPSKAABELHRVDLVLPEDTTEYFVATK 110
QY 64 LGNLTTIKHPSSEIIIGVLPEDYTRADEPGRGRRPGRPRKMPRHESSTLSMESPRKTM 123
DB 111 AG-----GVCFKPGTGMLERP-----PPGRPEKPEGANGSSARRPRYLL 151
QY 124 TDSKIMFELRGKPPREMIAGREBEYSIGRAMVK-----GHMNEYEPKAKQRTDYAPNL 178
DB 152 SAR-----ERTGR-----GARRKVECVCLPGW-----HGSPSC 180
QY 179 AVDYLAEREIHRMPRPDSIPELPIVPSRI-----DEPD-----KKGWCAHORRRRT 246
DB 181 GVPYV--QYSNLPFKERLVPR--EVPRRVAINAVNHSEFLDLVRPHELGDDVDAFVC 236
QY 213 -----ATVDP--YETDLKN--EYIRHMKOV-----KKGWCAHORRRRT 246
DB 237 ESNFTAYGEPRLPKFREMLTNGTEYIRH-KVLVYFLDHPFGGRQDGMADYLR 292

RESULT 3
US-08-849-281A-4
/ Sequence 4, Application US/08849281A
/ Patent No. 6153433
/ GENERAL INFORMATION:
/ APPLICANT: Eiichi MIYOSHI et al.

/ TITLE OF INVENTION: INHIBITOR FOR VIRAL REPLICATION
/ NUMBER OF SEQUENCES: 4
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
/ STREET: 2033 K Street, N.W., Suite 800
/ CITY: Washington
/ STATE: D.C.
/ COUNTRY: U.S.A.
/ ZIP: 20006
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: MS-DOS
/ SOFTWARE: Wordperfect 5.1
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/849,281A
/ FILING DATE: May 30, 1997
/ CLASSIFICATION: 514
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Lee Cheng
/ REGISTRATION NUMBER: 40,949
/ REFERENCE/DOCKET NUMBER: 97-0529+/LC (MC) /1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 202-721-8200
/ TELEFAX: 202-721-8250
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 4:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 531 amino acid residues
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: polypeptide
US-08-849-281A-4

Query Match 6.5%; Score 95.5; DB 3; Length 531;
Best Local Similarity 22.6%; Pred. No. 0.16;
Matches 67; Conservative 27; Mismatches 90; Indels 113; Gaps 17;
QY 6 PLAFLPLPDGDRNARNONDLISGGPL--PLESPSKLTSLSYDPTVPESDMKFAKR 63
DB 53 PVTPOASPPGPGDLRLT-PLYSHSPLQLPL-PPSKAABELHRVDLVLPEDTTEYFVATK 110
QY 64 LGNLTTIKHPSSEIIIGVLPEDYTRADEPGRGRRPGRPRKMPRHESSTLSMESPRKTM 123
DB 111 AG-----GVCFKPGTGMLERP-----PPGRPEKPEGANGSSARRPRYLL 151
QY 124 TDSKIMFELRGKPPREMIAGREBEYSIGRAMVK-----GHMNEYEPKAKQRTDYAPNL 178
DB 152 SAR-----ERTGR-----GARRKVECVCLPGW-----HGSPSC 180
QY 179 AVDYLAEREIHRMPRPDSIPELPIVPSRI-----DEPD-----KKGWCAHORRRRT 246
DB 181 GVPYV--QYSNLPFKERLVPR--EVPRRVAINAVNHSEFLDLVRPHELGDDVDAFVC 236
QY 213 -----ATVDP--YETDLKN--EYIRHMKOV-----KKGWCAHORRRRT 246
DB 237 ESNFTAYGEPRLPKFREMLTNGTEYIRH-KVLVYFLDHPFGGRQDGMADYLR 292

RESULT 4
US-09-252-991A-28311
/ Sequence 28311, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28311
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28311

Query Match 6.5%; Score 95; DB 4; Length 530;
Best Local Similarity 23.6%; Pred. No. 0.18;
Matches 65; Conservative 27; Mismatches 91; Indels 92; Gaps 16;

QY 15 DGDNRARNDPILSGGPIPLE-SPSRKLTSLSYDPTVESPDMKFAK-KRLGNL--- 68
DB 173 DG-RQRAPADVATAGLPFGHFGRR-----NRPD-QFLQHRRLPALPDR 218
QY 69 TTIKHSEIIGVLPEDYTRADEBPGRGRPPG---RPRKMRHESSTLSMSPKTKMT 124
DB 219 LDARHHPERQPGPVAGH-----PHRRRRRGPADPRRRGLPTLNREPTMLDERLIQE 272
QY 125 RDSKIMFELRGK-----PREMIAGRFEEYSLSGRAMV-----KGH----- 159
DB 273 AANRLDAERSQOVRQSLDYPDIA--IEDAYAIQRAWVERKIDGKVLKSHKIGLTSR 330
QY 160 ---MNNVEP-----IKAORTDPAVNLAVDYL-----ACREIH 189
DB 331 AMQVSNISEPDYGALDDMFEEGSDIIPQRF-IVPRVEVELATILKGLKPGCTLF 389
QY 190 RMPRPDKSIPELPYPSRIDEDATVDPRTDVK 224
DB 390 VLEATEWVIPALEIIDARIQ---VDPQTKVTRK 420

RESULT 5
US-09-252-991A-25361
; Sequence 25361, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25361
; LENGTH: 1037
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25361

Query Match 6.5%; Score 94.5; DB 4; Length 1037;
Best Local Similarity 21.5%; Pred. No. 0.55;
Matches 70; Conservative 36; Mismatches 103; Indels 117; Gaps 19;

QY 14 EDGDRNARNQ-----DPLISGGPLPLESPSRKLTSLSYDPTVESPDMKFAKRL 64
DB 161 EPGDRQRQSHGAMLPWRDP--BELPFPARATGRRLRR--SHPPAQEPQSSHQRLLR- 215
QY 65 GNLTTIKHHPSEIIGVLPEDYTRADERP---GRO-----GPPPG----- 101
DB 216 ---ORHQR---YRRSRKPAHRLARQIQRRPRQGDALQRPGGYLEPA 260
QY 102 RPRKMR-----HESSTLSMSPR---KTMTRDSKIMFELRGKPREMIAGR-----FE 146

DB 261 LPRPQAVAFQPIPDALALLODPBPGQPARBDRQPLQBRADP-----GRRRPVAE 315
QY 147 EYSLGRW---YGHMNNNEE-PIKAORTYANRLANDVYLACREIHMPPRDSIPELP 202
DB 316 PQAQRG-WRHPVRLHRAERDPLRLQR-----HAVRLQRRRLALVEDP 360
QY 203 I-----VPSRIDEFDATVDPRY---ETDLKNEYIRHWKQVKKGCAGHQRRTA 247
DB 361 VPGQGLPRRPGGLAVRRRLRWPGRBDRGRCQGRGLRPLQPOHDE-----RPA 409
QY 248 PHASISALINKIYQPGESKIVEQALG 273
DB 410 LGTRPVALLRLPRRRPGQPGQRPALG 435

RESULT 6
US-09-335-011-1
; Sequence 1, Application US/09335011
; Patent No. 6686450
; GENERAL INFORMATION:

; APPLICANT: Liu, Jun O
; APPLICANT: Sun, Luo
; TITLE OF INVENTION: IMMUNOSUPPRESSIVE AGENTS THAT INHIBIT CALCINEURIN
; FILE REFERENCE: 0492611-0345 (MT7280)
; CURRENT APPLICATION NUMBER: US/09/335,011
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2220
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-335-011-1

Query Match 6.3%; Score 91.5; DB 4; Length 2220;
Best Local Similarity 23.5%; Pred. No. 3.6; Length 2220;
Matches 76; Conservative 35; Mismatches 117; Indels 95; Gaps 19;

QY 13 PEDGDRNARNQ--NDPLISG-GPLPLESPSRKLTSLSYDPT---VPESPDMKFAK--- 62
DB 1696 PEDQEGELPQPKKPPPLADGSGPGF--EPGKV-GILNHRPVAMADGSDAGSGERKDES 1752
QY 63 -RLGNLTTIKHHPSEIIGVLPEDYTRADER-----PGRGR---PGRPRKMRHES 112
DB 1753 PRAG-----PTEPMDTSEATVCHSDLERTPPLRPGRPARDGSPSRPTLSLEILS 1803
QY 113 TSLM-----ESPRTKMTDRSKIMFELR-GKPFEMIA 142
DB 1804 ISARQOFTPLTPAQAPAPAPATTGTRAGHPPEBPLSRKRKRLDSTESGKTLILDA 1863
QY 143 GRFEE-----YSLGRAMVGHMNNNEEYPIKAORTDPAVNLAVDYLAEREIH-----RMP 192
DB 1864 YRWVQGGQGVAYVLGR--VERIMSETYMLKQVDEEALQAVXF--C-QVHGAALQR 1918
QY 193 RPKDSIPELPV--SRIDEFDATVDPRYETDLKNEYIRHWKQVKKGCAGHQRRTA 245
DB 1919 QASGDPTTPRGHPDPSRENFPVTVVPTAPDPVAD-----SVQRPSDATTKRPALAA 1972
QY 246 ---TAPHASISALINKIYQPG 263
DB 1973 ATTIIITCPASASASTLDSKDPG 1995

RESULT 7
US-09-248-796A-20111
; Sequence 20111, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Kelch Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN

FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 20111
LENGTH: 580
TYPE: PRT
ORGANISM: Candida albicans
FEATURE:
NAME/KEY: UNSURE
LOCATION: (19)
OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknown
US-09-248-796A-20111

Query Match 6.2%; Score 90.5; DB 4; Length 580;
Best Local Similarity 20.0%; Pred. No. 0.64;
Matches 41; Conservative 27; Mismatches 66; Indels 71; Gaps 7;

QY 6 PLAEPLPDGDRNARQNDPLISGGPLPSSPSRKLTSLSY----- 47
DB 376 PTDQFIPDNKR-----QSINELKLSLKYLTVMTPKNNLIFSTPR 419
QY 48 -----DPTVPSPPDMKFKARKLGNLTTIKHPSRIIGVLP----- 83
DB 420 NIRTTEPQMVNSDQKGTNETNNFNTPLHH-QQLIGSSPLKRRRESVSTSIISNT 478
QY 84 -EDYTRADEPGRQGRPPGRPRKMPRHESSTLSMESPRKMTTRDSKIMFELGKP-FEMI 141
DB 479 NNNHHFENRYKSGPKPDP-----VALESSPFIISLSDENLMGLRPKTLFDSI 529
QY 142 -AGREERYSLGRAWYKGMNNEYE 165
DB 530 TTSPQDKYDDEKVVYKNSHNNNEYD 554

RESULT 8

US-08-769-309A-5
Sequence 5, Application US/08/69309A
Patent No. 5741890
GENERAL INFORMATION:
APPLICANT: Scott, John D.,
APPLICANT: Naert, Brian J.,
APPLICANT: Klauck, Theresa M.
TITLE OF INVENTION: Protein Binding Domains of Gravin
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower/233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/769,309A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 5741890and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/33451
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1780 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-769-309A-5

Query Match 6.2%; Score 90; DB 1; Length 1780;
Best Local Similarity 22.6%; Pred. No. 3.7;
Matches 53; Conservative 32; Mismatches 93; Indels 56; Gaps 10;

QY 31 PLESPSRKLTSLSYDPTVPSPPDMKFKARKLGNLTTIKHPSRIIGVLPDPDYRAD 90
DB 663 PKP-BEPRKVDTSVSWALICVSSKKRRARRSS--SDEGGPKAWG---DHQKAD 714
QY 91 EE-----PGRQGRPPGRPRKMPRHESSTLSMESPRKMTTRDSKIMFE 132
DB 715 EAGDKETGTGILAGSGEHDPGSSSPEDAGSPTEGEGVSTWESFRLVTPRKSKSK 774
QY 133 LRKPFEMIAGRFEEYSLSGRAWYKGMNNEYEPIKAQRTDYAPNLAVDIACREI--H 189
DB 775 LEEKSEDSIAGSGVE-----HSTPTEPKER-----SWVSIRKFIPIGR 813
QY 190 RMPPDDKSIPLPLIV---PSRIDEPDATVD---PRYETDLKNEYIRHMKVKKG 237
DB 814 RKKPDKQEQAPVADAGPTGAMEDSDVPAVVLSEYDAVEREKMAQQAQKQK 867

RESULT 9

US-08-994-570-5
Sequence 5, Application US/08994570
Patent No. 6090929
GENERAL INFORMATION:
APPLICANT: Scott, John D.,
APPLICANT: Naert, Brian J.,
APPLICANT: Klauck, Theresa M.
TITLE OF INVENTION: Protein Binding Domains of Gravin
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower/233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/994,570
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: No. 6090929and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/33451
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1780 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-994-570-5

Query Match 6.2%; Score 90; DB 3; Length 1780;
Best Local Similarity 22.6%; Pred. No. 3.7;

Db 262 BEPLPTDSEVPNLBQLP-----ETELPDE 285

RESULT 13

US-09-270-767-33141
; Sequence 33141, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33141
; LENGTH: 773
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-33141

Query Match 6.1%; Score 89.5; DB 4; Length 773;

Best Local Similarity 19.8%; Pred. No. 1.2;

Matches 55; Conservative 38; Mismatches 80; Indels 105; Gaps 14;

QY 3 EIDPLAEFLP-----EDGDRNARNDPLISGGLPLSPSKRLTSLSYDPTVES 54

Db 157 DVDVLAECRQPMSEVSEHSEKRGDVNGHDAFGQADGGLPVEN-----LYLPDLDR 207

QY 55 P-----DMKFARKRLGNLLT---TIKHPSEIIGVL-----EDYT 87

Db 208 PCSALSERQEIKLVEBEIAAVLSGSEVYLKEH--NVLGIFPKPGVCMSSDYLRLSINEVT 265

QY 88 RADEPFGQGRP-----PGRPRKMPRHESSTLSMESPR-----KTMTRDSKI 129

Db 266 KTKLEKDKENQWSTFLQRPNRP--VPKSKQS---LEAERRAANAAYKVTIVKSAPREKSP 320

QY 130 MFEIARGKPFEMIAGRFEEEYSLGRAMVKGHNNEVEPIKAQRTDYAPVLAVDYLAACREIH 189

Db 321 MPEAKPAKKEATPPREVEK-----KEEVPVEEVEPEPEPEKO----- 359

QY 190 RMPPR-DKSIPELPIVPSRIDEPDATVDPRYETDLKNE 226

Db 360 BEPLPTDSEVPNLBQLP-----ETELPDE 383

RESULT 14

US-09-270-767-48358
; Sequence 48358, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 48358
; LENGTH: 773
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-48358

Query Match 6.1%; Score 89.5; DB 4; Length 773;

Best Local Similarity 19.8%; Pred. No. 1.2;

Matches 55; Conservative 38; Mismatches 80; Indels 105; Gaps 14;

QY 3 EIDPLAEFLP-----EDGDRNARNDPLISGGLPLSPSKRLTSLSYDPTVES 54

Db 157 DVDVLAECRQPMSEVSEHSEKRGDVNGHDAFGQADGGLPVEN-----LYLPDLDR 207

QY 55 P-----DMKFARKRLGNLLT---TIKHPSEIIGVL-----EDYT 87

Db 208 PCSALSERQEIKLVEBEIAAVLSGSEVYLKEH--NVLGIFPKPGVCMSSDYLRLSINEVT 265

QY 88 RADEPFGQGRP-----PGRPRKMPRHESSTLSMESPR-----KTMTRDSKI 129

Db 266 KTKLEKDKENQWSTFLQRPNRP--VPKSKQS---LEAERRAANAAYKVTIVKSAPREKSP 320

QY 130 MFEIARGKPFEMIAGRFEEEYSLGRAMVKGHNNEVEPIKAQRTDYAPVLAVDYLAACREIH 189

Db 321 MPEAKPAKKEATPPREVEK-----KEEVPVEEVEPEPEPEKO----- 359

QY 190 RMPPR-DKSIPELPIVPSRIDEPDATVDPRYETDLKNE 226

Db 360 BEPLPTDSEVPNLBQLP-----ETELPDE 383

RESULT 15

US-09-270-767-36557
; Sequence 36557, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36557
; LENGTH: 808
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-36557

Query Match 6.1%; Score 89.5; DB 4; Length 808;

Best Local Similarity 19.8%; Pred. No. 1.3;

Matches 55; Conservative 38; Mismatches 80; Indels 105; Gaps 14;

QY 3 EIDPLAEFLP-----EDGDRNARNDPLISGGLPLSPSKRLTSLSYDPTVES 54

Db 130 DVDVLAECRQPMSEVSEHSEKRGDVNGHDAFGQADGGLPVEN-----LYLPDLDR 180

QY 55 P-----DMKFARKRLGNLLT---TIKHPSEIIGVL-----EDYT 87

Db 181 PCSALSERQEIKLVEBEIAAVLSGSEVYLKEH--NVLGIFPKPGVCMSSDYLRLSINEVT 238

QY 88 RADEPFGQGRP-----PGRPRKMPRHESSTLSMESPR-----KTMTRDSKI 129

Db 239 KTKLEKDKENQWSTFLQRPNRP--VPKSKQS---LEAERRAANAAYKVTIVKSAPREKSP 293

QY 130 MFEIARGKPFEMIAGRFEEEYSLGRAMVKGHNNEVEPIKAQRTDYAPVLAVDYLAACREIH 189

Db 294 MPEAKPAKKEATPPREVEK-----KEEVPVEEVEPEPEPEKO----- 332

QY 190 RMPPR-DKSIPELPIVPSRIDEPDATVDPRYETDLKNE 226

Db 333 BEPLPTDSEVPNLBQLP-----ETELPDE 356

Search completed: April 9, 2005, 07:57:20
Job time : 27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2005, 07:56:26 ; Search time 336 Seconds
(without alignments)
271.724 Million cell updates/sec

Title: US-09-087-136-1

Perfect score: 1462
Sequence: 1 MSEIDPLAEFLPDPDGRNA.....INKTYGSEKTVQAGLI 275

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1418010 seqs, 331997259 residues

Total number of hits satisfying chosen parameters: 1418010

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1462	100.0	275	9	US-09-220-091-1
2	95.5	6.5	448	14	US-10-156-761-9046
3	95.5	6.5	533	14	US-10-029-386-34087
4	94.5	6.5	410	14	US-10-080-170-216
5	94.5	6.5	410	16	US-10-080-170-216
6	94.5	6.5	410	16	US-10-468-356-216
7	93.5	6.4	383	9	US-09-738-692-4870
8	93	6.4	413	15	US-10-424-599-184643
9	92	6.3	204	16	US-10-437-863-138027
10	90	6.2	1781	9	US-09-738-877-3
11	90	6.2	1781	10	US-09-961-403-13
12	90	6.2	1781	15	US-10-428-487-16
13	90	6.2	1781	15	US-10-211-462-44

14	89.5	6.1	845	14	US-10-205-219-133	Sequence 133, App
15	89.5	6.1	847	14	US-10-153-668-450	Sequence 450, App
16	89	6.1	331	15	US-10-094-749-2877	Sequence 2877, App
17	89	6.1	530	14	US-10-156-761-9582	Sequence 9582, App
18	88.5	6.1	399	14	US-10-314-657-37	Sequence 37, App1
19	88.5	6.1	1704	16	US-10-408-765A-2220	Sequence 2220, App
20	88.5	6.1	1704	16	US-10-648-593-213	Sequence 43259, App
21	88	6.0	439	15	US-10-425-114-43259	Sequence 120, App
22	87.5	6.0	1704	14	US-10-340-097-120	Sequence 120, App
23	87.5	6.0	1704	14	US-10-336-215-120	Sequence 120, App
24	87.5	6.0	1704	14	US-10-336-215-120	Sequence 120, App
25	87	6.0	303	14	US-10-156-761-12125	Sequence 12125, App
26	87	6.0	332	16	US-10-437-963-128558	Sequence 128558, App
27	87	6.0	785	9	US-09-801-368-122	Sequence 122, App
28	87	6.0	4952	15	US-10-051-874-56	Sequence 56, App
29	87	6.0	5008	15	US-10-051-874-166	Sequence 166, App
30	87	6.0	5159	15	US-10-085-198-112	Sequence 112, App
31	87	6.0	5262	15	US-10-051-874-165	Sequence 165, App
32	87	6.0	5262	15	US-10-051-874-167	Sequence 167, App
33	86.5	5.9	645	13	US-10-096-241-10	Sequence 93, App1
34	86.5	5.9	645	14	US-10-082-747A-93	Sequence 93, App1
35	86.5	5.9	645	14	US-10-207-498-4	Sequence 4, App11
36	86.5	5.9	645	15	US-10-609-370-3	Sequence 3, App11
37	86.5	5.9	669	9	US-09-773-517-1	Sequence 1, App11
38	86.5	5.9	669	9	US-09-773-517-1	Sequence 1, App11
39	86.5	5.9	669	9	US-09-849-868-1	Sequence 1, App11
40	86.5	5.9	669	14	US-10-022-609-11	Sequence 11, App1
41	86.5	5.9	669	14	US-10-453-883-1	Sequence 1, App1
42	86.5	5.9	675	9	US-09-773-517-3	Sequence 3, App11
43	86.5	5.9	675	9	US-09-792-025-3	Sequence 3, App11
44	86.5	5.9	675	9	US-09-849-868-3	Sequence 3, App11
45	86.5	5.9	675	14	US-10-453-183-3	Sequence 3, App11

ALIGNMENTS

RESULT 1

US-09-220-091-1

Sequence 1, Application US/09220091

Patent NO. US20020064523A1

GENERAL INFORMATION:

APPLICANT: H. Robert Horvitz

APPLICANT: Craig Ceol

APPLICANT: Xiaowei Lu

TITLE OF INVENTION: A TUMOR SUPPRESSOR PATHWAY IN C. ELEGANS

FILE REFERENCE: 01997/202003

CURRENT APPLICATION NUMBER: US/09/220,091

CURRENT FILING DATE: 1998-12-23

EARLIER APPLICATION NUMBER: 60/047,996

EARLIER FILING DATE: 1997-05-28

EARLIER APPLICATION NUMBER: 09/087,136

EARLIER FILING DATE: 1998-05-28

NUMBER OF SEQ ID NOS: 19

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 1

LENGTH: 275

TYPE: PRT

ORGANISM: Caenorhabditis elegans

US-09-220-091-1

Query Match 100.0%; Score 1462; DB 9; Length 275;

Best Local Similarity 100.0%; Pred. No. 1.2e-128;

Matches 275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSEIDPLAEFLPDPDGRNAQNDPLISGGLPESPRKLTSLSYDPTVPESDMFA 60

Db 1 MSEIDPLAEFLPDPDGRNAQNDPLISGGLPESPRKLTSLSYDPTVPESDMFA 60

Qy 61 RKRGLNLTITIKHPSEITIGVLPEDYTRADEEPGRQGRPPGRPRKMPHESSTLMESPR 120

Db 61 RKRGLNLTITIKHPSEITIGVLPEDYTRADEEPGRQGRPPGRPRKMPHESSTLMESPR 120

Db 258 TSEPLV---EPWQSSQSSSEQADFDAGSPRRSRQPTQSDPELRSQPREIRDA 313
Qy 128 KIMFELRCKPFEMINGRFESEYSLGRAWVKGMNNEYEPKQRTDYAPNLAVDIACRE 187
Db 314 ---YGRSGPYE-----WPTNH-SSHLEPYRRYKQGPPEHTEHGQLYE 354
Qy 188 IHRMPR---PDKSIPELPIVPSRIDEPDATVDPYE 220
Db 355 RYKQPRRATPPRASVNPISQVRYRGSTARDPRVD 391

RESULT 5
US-10-080-170-216

Sequence 216, Application US/10080170
Publication No. US20040121322A9
GENERAL INFORMATION:
APPLICANT: COLE, S.T.
TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
TREATMENT OF MYCOBACTERIOSES
FILE REFERENCE: 03495.0218
CURRENT APPLICATION NUMBER: US/10/080,170
CURRENT FILING DATE: 2002-06-10
PRIOR APPLICATION NUMBER: 60/270,123
PRIOR FILING DATE: 2001-02-22
NUMBER OF SEQ ID NOS: 652
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 216
LENGTH: 410
TYPE: PRT
ORGANISM: Mycobacterium leprae
US-10-080-170-216

Query Match 6.5%; Score 94.5; DB 16; Length 410;
Best Local Similarity 20.3%; Pred. No. 3.9;
Matches 44; Conservative 30; Mismatches 94; Indels 49; Gaps 8;

Qy 17 DRNARQNDPLISGGLPLESPSRKLTSLSYDP---TVPESDPMKFAKRLGNLTTIKH 73
Db 211 DQNHANQDTAD-----TELMHSDPPLRTQRTADNRSARTHSQARVVEY 257
Qy 74 HPSEIIQVLPEDYTRADEPGRQ-----GRPGPRKPKRHSSTSLMESPRKMTDTS 127
Db 258 TSEPLV---EPWQSSQSSSEQADFDAGSPRRSRQPTQSDPELRSQPREIRDA 313
Qy 128 KIMFELRCKPFEMINGRFESEYSLGRAWVKGMNNEYEPKQRTDYAPNLAVDIACRE 187
Db 314 ---YGRSGPYE-----WPTNH-SSHLEPYRRYKQGPPEHTEHGQLYE 354
Qy 188 IHRMPR---PDKSIPELPIVPSRIDEPDATVDPYE 220
Db 355 RYKQPRRATPPRASVNPISQVRYRGSTARDPRVD 391

RESULT 6
US-10-468-356-216

Sequence 216, Application US/10468356
Publication No. US20040197896A1
GENERAL INFORMATION:
APPLICANT: COLE, STEWART
TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
TREATMENT OF MYCOBACTERIOSES
FILE REFERENCE: 05394.0019
CURRENT APPLICATION NUMBER: US/10/468,356
CURRENT FILING DATE: 2003-08-19
PRIOR APPLICATION NUMBER: 10/080,170
PRIOR FILING DATE: 2002-02-22
PRIOR APPLICATION NUMBER: 60/270,123
PRIOR FILING DATE: 2001-02-22
NUMBER OF SEQ ID NOS: 655
SOFTWARE: PatentIn Ver. 3.2

SEQ ID NO 216
LENGTH: 410
TYPE: PRT
ORGANISM: Mycobacterium leprae
US-10-468-356-216

Query Match 6.5%; Score 94.5; DB 16; Length 410;
Best Local Similarity 20.3%; Pred. No. 3.9;
Matches 44; Conservative 30; Mismatches 94; Indels 49; Gaps 8;

Qy 17 DRNARQNDPLISGGLPLESPSRKLTSLSYDP---TVPESDPMKFAKRLGNLTTIKH 73
Db 211 DQNHANQDTAD-----TELMHSDPPLRTQRTADNRSARTHSQARVVEY 257
Qy 74 HPSEIIQVLPEDYTRADEPGRQ-----GRPGPRKPKRHSSTSLMESPRKMTDTS 127
Db 258 TSEPLV---EPWQSSQSSSEQADFDAGSPRRSRQPTQSDPELRSQPREIRDA 313
Qy 128 KIMFELRCKPFEMINGRFESEYSLGRAWVKGMNNEYEPKQRTDYAPNLAVDIACRE 187
Db 314 ---YGRSGPYE-----WPTNH-SSHLEPYRRYKQGPPEHTEHGQLYE 354
Qy 188 IHRMPR---PDKSIPELPIVPSRIDEPDATVDPYE 220
Db 355 RYKQPRRATPPRASVNPISQVRYRGSTARDPRVD 391

RESULT 7

US-09-738-626-4870
Sequence 4870, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:

APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, MAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 4870
LENGTH: 383
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-4870

Query Match 6.4%; Score 93.5; DB 9; Length 383;
Best Local Similarity 21.0%; Pred. No. 4.4; Indels 115; Gaps 17;
Matches 71; Conservative 49; Mismatches 103;

Qy 2 SEIDPLAEFLPE-----DGRNARQNDPLISGGLP-----LESFPRKLTSLSYDP 49
Db 82 SMVDGLVSAIPBALQNDGPQSKR---MLSGVAIPGATVTAADPTTK-----DP 130
Qy 50 TVPESPDMKFAKRLGNLTTIKHPSIIIGVLP-----DYTRADEE 92
Db 131 T-----PTL-----ISRLEIATRLQKHGSGILITLDELQGANVDLHVLAIVADLRDPDF 183
Qy 93 PG--RQGRPPGRPKRPHHSSTSLMESPRKMT-----RDSKIM----- 131

```
Db      184  IALVAAGLPBEGIDRL-QHEGTFTRRARIILNFVNHEDSVEMFLDTRABSGQRMTSEA 242
      132  -EL-----RGKFEMIAGRFEEYSIGRAWVGHNNNEPERIKARTYAPNALAVDYIAC 185
      243  AELAAQISKGYYSM-----QLNGSLAMASTLDNS-DTIGAPQVAVRDEVVRRGM 294
Qy      186  R-----EIRMPRPDSISELPVPSRIDEFATVDPRTYETDLKNEYIKHWQYKGGWCAH 241
      295  QVHEPSLHQVDPGBELTILYALIAQLSKNGEMVSTGD-----IAHLMGKKNALSM 343
Qy      242  QRR-----TAPARSIALLNKITYOP 262
      344  QKQILSRGLVEVPKKGFLNFTLPYMRHLLNSPHHRP 381
Db

RESULT 8
US-10-424-599-184643
; Sequence 184643, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 184643
; LENGTH: 413
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(413)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_137749C.1.pdp
US-10-424-599-184643

Query Match
Best Local Similarity 22.4%; Score 93; DB 15; Length 413;
Matches 66; Conservative 36; Mismatches 91; Indels 102; Gaps 15;

Qy      31  PLPLESPSR-----KLTSLSYDPTVP-----ESPDMKFAKKRLGNLTTIKHHPSEII 79
      105  PLPVIAQAKPRGRPRPKLKSQPTLWPAALNDNPALQSAKRGPG-----RPKKIA 156
Db      80  GVLPEDYTRADEEPFGQRPFG-----RPRKPRHSSSTSLMES-----PRKT 122
      157  G--PVGVSPPMPVPGRRPGRTGSKLPRKPRGRPKSYSAISSGLKRRXPGRPVPA 214
Qy      123  WTRDSKIMEL-----RGKPFMIAGRFEESY-----L 151
      215  ESNVANITPPRAAVAPGLPTVQPIVTPASVPNGSPRRKGPCKIIVAGAGAPALSSVGAGR 274
Db      152  GRAWVGHNNNEPERIKARTYAPNALAVDYIACREIHHMPPDKSIPPLPIVPSRIDEF 211
      275  GRGRPRG-----VLPLVRPGRPQKLAVG-----RP-KMPARRPVAPQCIDAI 315
Qy      212  DATVDPRTYETDLKNEYIKHWK-QYK-----KGWCAHQRRRTAPARSIALLNKI 259
      316  --TAHRAANDDLRXKRLHEHFQSKVKSIGCTLKYEFNHESPVA-----IAAIQEL 363
Db

RESULT 9
US-10-437-963-138027
; Sequence 138027, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrew A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 138027
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(204)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_39453C.1.pdp
US-10-437-963-138027

Query Match
Best Local Similarity 30.5%; Score 92; DB 16; Length 204;
Matches 32; Conservative 12; Mismatches 47; Indels 14; Gaps 4;

Qy      15  DGRNARQNDPLISGPPLESPSRKLTSLSYDPTVPESPDMPKARRKLNLTITIKH 74
      58  DGRSRSSTSPYS--PIPLKTPDPKASSVHRGARVG-----CRARVG--SNGGCH 105
Db      75  PSEIIGVLPEDYTRADEEPFGQRPGRPR--RKPFRHSSSTSLME 117
      106  NSNAGTVPVPAVPVPASSSSVSMGKPSRRPRRRRWMLHSEAVVME 150
Qy

RESULT 10
US-09-738-877-3
; Sequence 3, Application US/09738877
; Patent No. US20020015970A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan
; APPLICANT: Weiss, Stephen J.
; APPLICANT: Glynn, Richard
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF ANGIOGENESIS, COMPOSITIONS, AND MET
; TITLE OF INVENTION: SCREENING FOR ANGIOGENESIS MODULATORS
; FILE REFERENCE: A-69806/DJB/JJD
; CURRENT APPLICATION NUMBER: US/09/738,877
; CURRENT FILING DATE: 2000-12-15
; PRIOR APPLICATION NUMBER: US 09/737,977
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: US 60/148,425
; PRIOR FILING DATE: 1999-08-11
; PRIOR APPLICATION NUMBER: PCT/US 00/22061
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1781
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-738-877-3

Query Match
Best Local Similarity 22.6%; Score 90; DB 9; Length 1781;
Matches 53; Conservative 32; Mismatches 93; Indels 56; Gaps 10;

Qy      31  PLPLESPSRKLTSLSYPTVPESPDMPKARRKLNLTITIKHHPSEIIGVLPEDYTRAD 90
```



```
Db 663 PKP-BEPKRVDTSVSWELICVGSKKRARRSS---SDEEGGPKAMG---DHQXAD 714
Qy 91 EE-----PGRQRPGRPRKMPRHSS-----TSMSPRKTMTDTSKIME 132
Db 715 EAGKDKETGTGDIAGSQEHDPGQSSSPGQAGSPTEGEGVSTWESFKRLVTPRKKSXK 774
Qy 133 LRGKPFEMIAGRFEEEYSIGRAMVKGHMNNEYEPKQRTDYAPMLADYLACREI---H 189
Db 775 LEKSEDSIAGSGVE-----HSTPDTEPKER-----SWSIKKFIPGR 813
Qy 190 RMPRPDKSIPELPIY---PSRIDEFATVD---PRYEDLKNBYIRHMKQVKKG 237
Db 814 RKRKPDGKQEQAPVEDAGPTGANEDSDVPVAVPLSEYDAVEREKMEAOQAQKG 867

RESULT 11
US-09-961-403-13
; Sequence 13, Application US/09961403
; Publication No. US2003007589A1
; GENERAL INFORMATION:
; APPLICANT: HE-STUMPF, HOLGER
; APPLICANT: HAENDLER, BERNARD
; APPLICANT: KRAETZSCHMAR, JOERN
; APPLICANT: KREFT, BERTHOLT
; APPLICANT: WINTERHAGER, ELKE
; APPLICANT: REGIDOR, PEDRO
; APPLICANT: SCOTTI, SIMONE
; TITLE OF INVENTION: METHOD FOR IN VITRO DIAGNOSIS OF ENDOMETRIOSIS
; FILE REFERENCE: SCH-1789
; CURRENT APPLICATION NUMBER: US/09/961,403
; CURRENT FILING DATE: 2001-09-25
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1781
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-961-403-13

Query Match 6.2%; Score 90; DB 10; Length 1781;
Best Local Similarity 22.6%; Pred. No. 75;
Matches 53; Conservative 32; Mismatches 93; Indels 56; Gaps 10;

Qy 31 PPLESPSRKLTSLSYDPTVPESPDMPKARRLGNLTTIGHSEIIGVLPEDYTRAD 90
Db 663 PKP-BEPKRVDTSVSWELICVGSKKRARRSS---SDEEGGPKAMG---DHQXAD 714
Qy 91 EE-----PGRQRPGRPRKMPRHSS-----TSMSPRKTMTDTSKIME 132
Db 715 EAGKDKETGTGDIAGSQEHDPGQSSSPGQAGSPTEGEGVSTWESFKRLVTPRKKSXK 774
Qy 133 LRGKPFEMIAGRFEEEYSIGRAMVKGHMNNEYEPKQRTDYAPMLADYLACREI---H 189
Db 775 LEKSEDSIAGSGVE-----HSTPDTEPKER-----SWSIKKFIPGR 813
Qy 190 RMPRPDKSIPELPIY---PSRIDEFATVD---PRYEDLKNBYIRHMKQVKKG 237
Db 814 RKRKPDGKQEQAPVEDAGPTGANEDSDVPVAVPLSEYDAVEREKMEAOQAQKG 867

RESULT 12
US-10-428-487-16
; Sequence 16, Application US/10428487
; Publication No. US20040006780A1
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA K.
; APPLICANT: GERBER, HANS-PETER
; TITLE OF INVENTION: VEGF-MODULATED GENES AND METHODS EMPLOYING THEM
; FILE REFERENCE: 09800080-0103
; CURRENT APPLICATION NUMBER: US/10/428,487
; CURRENT FILING DATE: 2003-05-02
; PRIOR APPLICATION NUMBER: 09/815,153
; PRIOR FILING DATE: 2001-03-21
```

```
; PRIOR APPLICATION NUMBER: 60/191,201
; PRIOR FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 1781
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-428-487-16

Query Match 6.2%; Score 90; DB 15; Length 1781;
Best Local Similarity 22.6%; Pred. No. 75;
Matches 53; Conservative 32; Mismatches 93; Indels 56; Gaps 10;

Qy 31 PPLESPSRKLTSLSYDPTVPESPDMPKARRLGNLTTIGHSEIIGVLPEDYTRAD 90
Db 663 PKP-BEPKRVDTSVSWELICVGSKKRARRSS---SDEEGGPKAMG---DHQXAD 714
Qy 91 EE-----PGRQRPGRPRKMPRHSS-----TSMSPRKTMTDTSKIME 132
Db 715 EAGKDKETGTGDIAGSQEHDPGQSSSPGQAGSPTEGEGVSTWESFKRLVTPRKKSXK 774
Qy 133 LRGKPFEMIAGRFEEEYSIGRAMVKGHMNNEYEPKQRTDYAPMLADYLACREI---H 189
Db 775 LEKSEDSIAGSGVE-----HSTPDTEPKER-----SWSIKKFIPGR 813
Qy 190 RMPRPDKSIPELPIY---PSRIDEFATVD---PRYEDLKNBYIRHMKQVKKG 237
Db 814 RKRKPDGKQEQAPVEDAGPTGANEDSDVPVAVPLSEYDAVEREKMEAOQAQKG 867

RESULT 13
US-10-211-462-44
; Sequence 44, Application US/10211462
; Publication No. US2004003495A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Aziz, Natsaba
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Angiogenesis, Compositions and
; FILE REFERENCE: 018501-006200US
; CURRENT APPLICATION NUMBER: US/10/211,462
; CURRENT FILING DATE: 2003-02-13
; PRIOR APPLICATION NUMBER: US 09/784,356
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: US 09/791,390
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: US 60/310,025
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/334,244
; PRIOR FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 230
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 44
; LENGTH: 1781
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-211-462-44

Query Match 6.2%; Score 90; DB 15; Length 1781;
Best Local Similarity 22.6%; Pred. No. 75;
Matches 53; Conservative 32; Mismatches 93; Indels 56; Gaps 10;

Qy 31 PPLESPSRKLTSLSYDPTVPESPDMPKARRLGNLTTIGHSEIIGVLPEDYTRAD 90
Db 663 PKP-BEPKRVDTSVSWELICVGSKKRARRSS---SDEEGGPKAMG---DHQXAD 714
Qy 91 EE-----PGRQRPGRPRKMPRHSS-----TSMSPRKTMTDTSKIME 132
Db 715 EAGKDKETGTGDIAGSQEHDPGQSSSPGQAGSPTEGEGVSTWESFKRLVTPRKKSXK 774
```

QY 133 LRGRFEMIAGRFEFEYSIGRAWVKGMMNEVEPIKAQRTDYAPMLAVDYLAICREI--H 189
DB 775 LEEKSEDSISAGV-----HSTPTEPEKE-----SWSIKKFIFGR 813
QY 190 RMPRPDKSIPELPIV---PSRIDEFPAVD---PREYTDLKNEXIRHMKQVKKG 237
DB 814 RKKRDKGQEQAPVEDAGFTGANNEDSDVPANVPLSEYDAVEREMEQQAQKG 867

RESULT 14
US-10-205-219-133
; Sequence 133, Application US/10205219
; Publication No. US20030138803A1
; GENERAL INFORMATION:
; APPLICANT: Warner-Lambert Company
; APPLICANT: Lee, Kevin
; APPLICANT: Dixon, Alister
; APPLICANT: Brooksbank, Robert
; APPLICANT: Pinnock, Robert
; TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain
; FILE REFERENCE: WL-A-018200
; CURRENT APPLICATION NUMBER: US/10/205,219
; CURRENT FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: GB 0118354.0
; PRIOR FILING DATE: 2001-07-27
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 133
; LENGTH: 845
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Matrin 3
US-10-205-219-133

Query Match 6.1%; Score 89.5; DB 14; Length 845;
Best Local Similarity 21.9%; Pred. No. 30;
Matches 63; Conservative 27; Mismatches 91; Indels 107; Gaps 14;

QY 28 SGGPLPLES-----PSRKLTSLSYDPTVPBSPDMKFAKRLGNLT 69
DB 91 SRGPLPLSQHRGADQASNLASFGLSARDLDELRY-----PEDKITPENLPQILL 143
QY 70 TIKHHPSEIIGVLPEDYTRADEEP---GROGRPPGR--PRKMPRHESSTLMSPRXTM 123
DB 144 QLKRR-----RTEGPTLSYGRDGRSATREPPYRVPRDD-----WEKKHF 184
QY 124 TRDSKIMFELRGKPEMTAGRFEEYSIG-RAWVKGMMNEVEPIKAQRTDYAPMLAVDY 182
DB 185 RDDS---FDDRGPSINPVL-----DYHGSRSQESGYD-----RMDYEDDLRDG 227
QY 183 LACR-----EIHMPRPDKSIPELPIV-----PSRIDEFPAVD 217
DB 228 ERCRDSSFGETSHNYHKPDSYERMGKGPPLQERSLFEKKRGAPPSNIEDPFGLLPX 287
QY 218 RYE-----TDLKNEXIRHMKQVKKGCAHQRRTPAHPARSIALINKIY 260
DB 288 GYPHLCSTCDLPVHSNKESQHING-----ASHSRRCQLLEIY 326

RESULT 15
US-10-153-668-450
; Sequence 450, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668

; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 450
; LENGTH: 847
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-153-668-450

Query Match 6.1%; Score 89.5; DB 14; Length 847;
Best Local Similarity 21.9%; Pred. No. 30;
Matches 63; Conservative 27; Mismatches 91; Indels 107; Gaps 14;

QY 28 SGGPLPLES-----PSRKLTSLSYDPTVPBSPDMKFAKRLGNLT 69
DB 91 SRGPLPLSQHRGADQASNLASFGLSARDLDELRY-----PEDKITPENLPQILL 143
QY 70 TIKHHPSEIIGVLPEDYTRADEEP---GROGRPPGR--PRKMPRHESSTLMSPRXTM 123
DB 144 QLKRR-----RTEGPTLSYGRDGRSATREPPYRVPRDD-----WEKKHF 184
QY 124 TRDSKIMFELRGKPEMTAGRFEEYSIG-RAWVKGMMNEVEPIKAQRTDYAPMLAVDY 182
DB 185 RDDS---FDDRGPSINPVL-----DYHGSRSQESGYD-----RMDYEDDLRDG 227
QY 183 LACR-----EIHMPRPDKSIPELPIV-----PSRIDEFPAVD 217
DB 228 ERCRDSSFGETSHNYHKPDSYERMGKGPPLQERSLFEKKRGAPPSNIEDPFGLLPX 287
QY 218 RYE-----TDLKNEXIRHMKQVKKGCAHQRRTPAHPARSIALINKIY 260
DB 288 GYPHLCSTCDLPVHSNKESQHING-----ASHSRRCQLLEIY 326

Search completed: April 9, 2005, 08:07:46
Job time : 337 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 9, 2005, 06:18:21 ; Search time 81 Seconds
(without alignments)
1313.075 Million cell updates/sec

Title: US-09-087-136-1
1462
Perfect score: 1 MSEIDPLAEFLPEDGDNA.....INKIYQGESKTVQALGLI 275
Sequence:

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1980s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1453	99.4	275	2	AAW83386
2	109	7.5	332	3	AAQ22132
3	109	7.5	332	3	AAQ40321
4	100	6.8	748	7	ABO69681
5	96	6.6	1262	8	ADQ97815
6	95.5	6.5	531	2	AAK48994
7	95.5	6.5	531	2	AAK97615
8	95.5	6.5	531	2	AAK97615
9	95.5	6.5	533	7	ADK55801
10	95.5	6.5	533	8	ABO60453
11	95.5	6.5	544	7	ADK68089
12	95	6.5	530	7	ABO79565
13	95	6.5	1161	4	ABR70419
14	94.5	6.5	410	5	ABU05565
15	94.5	6.5	1037	7	ABO76615
16	93.5	6.4	192	3	AAQ40322
17	93.5	6.4	192	3	AAQ22133
18	93.5	6.4	383	4	AAQ91116
19	93.5	6.4	531	2	AAW24015
20	93	6.4	325	4	ABK69792
21	93	6.4	381	4	ABK59442
22	92.5	6.3	809	4	ABG09518
23	92.5	6.3	1359	8	ADP75957
24	92	6.3	491	7	ADP69374
25	92	6.3	520	8	ADG65858

26	92	6.3	535	7	ADD68093	ADD68093 Human Gnt
27	91.5	6.3	1977	6	AAE37958	AAE37958 Human kin
28	91.5	6.3	2191	6	AAE37965	AAE37965 Human ORF
29	91.5	6.3	2220	3	ABK42235	ABK42235 Human ORF
30	91.5	6.3	2220	3	AAV54004	AAV54004 Full leng
31	91.5	6.3	2220	7	ADK55126	ADK55126 Human Pro
32	91.5	6.3	2220	7	ADK55130	ADK55130 Human Pro
33	91.5	6.3	2421	4	ABK65643	ABK65643 Drosophil
34	91	6.2	273	3	AAQ31861	AAQ31861 Arabidops
35	91	6.2	276	3	AAQ31860	AAQ31860 Arabidops
36	91	6.2	309	3	AAQ31859	AAQ31859 Arabidops
37	90.5	6.2	246	6	ABP60672	ABP60672 Beta tran
38	90	6.2	233	6	ABM72432	ABM72432 Staphyloc
39	90	6.2	1780	2	AAW53863	AAW53863 Human gra
40	90	6.2	1780	2	AAW53860	AAW53860 Human gra
41	90	6.2	1781	5	AAO17365	AAO17365 Human gra
42	90	6.2	1781	5	ABU03477	ABU03477 Angiogene
43	90	6.2	1783	5	ABK97448	ABK97448 Novel hum
44	89.5	6.1	216	4	ABG06379	ABG06379 Novel hum
45	89.5	6.1	339	7	ADG77007	ADG77007 Human nuc

ALIGNMENTS

RESULT 1
ID AAW83386 standard; protein, 275 AA.
XX
AC AAW83386;
DT 29-MAR-1999 (first entry)
XX
DE Caenorhabditis elegans symMuv protein LIN-37.
XX
KW LIN-37; synthetic multivulvar; symMuv; signal transduction; animal model;
KW tumour suppressor; retinoblastoma; cancer; cancer; cell proliferation;
KW gene therapy.
XX
OS Caenorhabditis elegans.
XX
PN WO9854299-A1.
XX
PD 03-DEC-1998.
XX
PF 28-MAY-1998; 98WO-US011043.
XX
PR 28-MAY-1997; 97US-0047996P.
XX
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
XX
PI Horvitz HR, Ceol C, Lu X;
XX
DR WPI, 1999-045362/04.
XX
PT Novel LIN-37, -35, -55, -52, -53 and -54 gene from C. elegans - useful
PT for treating diseases associated with altered levels of cell
PT proliferation, e.g. carcinomae.
PS Claim 7; Fig 2; 70pp; English.
XX
CC This is the amino acid sequence of LIN-37, a novel 32 kDa hydrophilic
CC protein of Caenorhabditis elegans. lin-37 cDNA (see AAV72859) was
CC isolated from the Okkema embryonic cDNA library. The lin-37 transcript is
CC about 1 kb in size and is present both in embryonic and mixed-stage RNAs.
CC The lin-35 gene is a novel synthetic multivulvar (symMuv) gene involved
CC in cell fate and cell proliferation, and is part of a pathway that may be
CC used as a genetic and biochemical model system for tumour suppression and
CC cancer in mammals. SymMuv pathway genes and proteins may be used to
CC identify genes which are part of the mammalian pathway and to identify
CC genes, proteins and therapeutic compounds which modulate this pathway.
CC Pure nucleic acids (see AAV72859-65) encoding C. elegans symMuv

CC polypeptides selected from LIN-37, -35, -55, -52, -53, -54 and E2F-1 (see
CC AAM8386-92), are new. Also claimed are: (1) nucleic acids (see AAV2866-
CC 67) encoding the mouse and human homologues of C. elegans LIN-54; (2)
CC vectors containing the nucleic acids; (3) transgenic cells; (4) a pure
CC mammalian SynMuv polypeptide, and (5) an antibody which binds to a SynMuv
CC family protein. The SynMuv nucleic acids and polypeptides can be used to
CC diagnose and treat, especially by gene therapy, conditions involving
CC altered levels of cell proliferation, e.g. SynMuv-associated carcinomas
XX

SO Sequence 275 AA;

Query Match 99.4%; Score 1453; DB 2; Length 275;
Best Local Similarity 99.6%; Pred. No. 2e-143;
Matches 274; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSEIDPLAFLLPEDGDRNARQNDPLISGGLPLESPSRKLTSLSYDPTVESPDMKFA 60
DB 1 MSEIDPLAFLLPEDGDRNARQNDPLISGGLPLESPSRKLTSLSYDPTVESPDMKFA 60
QY 61 RKRIGNLTTIKHPEEIIIGVLPEDYTRADEEGRGPRGRPRKMPHRESSTIMESEPR 120
DB 61 RKRIGNLTTIKHPEEIIIGVLPEDYTRADEEGRGPRGRPRKMPHRESSTIMESEPR 120
QY 121 KMTNRDSKIMFELRGKPFEMIAGRFEEYSLSRAWVKGHNNNEYEPKAKQRTDYAPNLAV 180
DB 121 KMTNRDSKIMFELRGKPFEMIAGRFEEYSLSRAWVKGHNNNEYEPKAKQRTDYAPNLAV 180
QY 181 DYLAEREIHRMPRPDKSIPELIVPSRIIDEPATYDPRYETDLKNEYIRHMKQVKKGWCA 240
DB 181 DYLAEREIHRMPRPDKSIPELIVPSRIIDEPATYDPRYETDLKNEYIRHMKQVKKGWCA 240
QY 241 HORRRTPARASIALINKIYOGESKTVQALGLI 275
DB 241 HORRRTPARASIALINKIYOGESKTVQALGLI 275

RESULT 2
AAG22132
ID AAG22132 standard; protein; 332 AA.
XX
AC AAG22132;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 24944.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX
XX Arabidopsis thaliana.
OS
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.

PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134376P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139482P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142820P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.

PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149923P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.

PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 22-OCT-1999; 99US-0160992P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161928P.
PR 28-OCT-1999; 99US-0161932P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 7.5%; Score 109; DB 3; Length 332;
Best Local Similarity 24.0%; Pred. No. 0.022;
Matches 69; Conservative 37; Mismatches 99; Indels 82; Gaps 16;

QY 5 DP---LAFFLPEDGDRVARQ-----NDPLISGGPLPESPSSKUTLSLSYPTVPEESP 56
DB 86 DPSAVLYPFALPGRC-FSARVVRGVADPSVTAGLSGYPPRPST-----YDPGEYEQRQ 140
QY 57 M-----KEPRKRLGNLTTIKHHPSELIGVLPEDYTRADEBGRGRRPG----- 101
DB 141 MESLQQFTIRERNPDI-----RPLPRGL-----GSPVGLGPIRASPOF 179
QY 102 -RPRKMPHBSSTLSMESP-RKMTWRSKIMFELRGKPFEMIAGRFEEBSLGAAMXG 158
DB 180 LQPRVAP---PPTSLDTSRNRKARSXDALAV-VRGKRVITEGS-SLYSLGRSWDK- 233
QY 159 HMNNEYEPKQRTDYNLAVDYLACREIHRMPRP-----DKSIPELPIVPSRID-- 209
DB 234 --NGAHVGIQPRSGI-----MKPLPKPLVDLTTETSVDDPDEBSADEBK 278
QY 210 EPDAIVDPREYETDLKNEYIRMKQYKKGCAHQRRRTAPHARSIALI 256
DB 279 EDEAVKQLSEKDLLKRIERAKKVRADLRERSRIRIRYKERTILI 325

RESULT 3
AAG40321
ID AAG40321 standard; protein; 332 AA.

AC AAG40321;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 50015.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-00301439.

PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0128845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138647P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0144055P.
PR 06-JUL-1999; 99US-0143390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.

PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 20-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 21-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 22-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149922P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150844P.
PR 27-AUG-1999; 99US-0151084P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151930P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154039P.
PR 16-SEP-1999; 99US-0154038P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155466P.
PR 24-SEP-1999; 99US-015659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.

OS Mus musculus.
 XX WO2004060304-A2.
 XX 22-JUL-2004.
 PD 22-DEC-2003; 2003WO-US041389.
 XX 27-DEC-2002; 2002US-00330773.
 XX (SAGR-) SAGRES DISCOVERY INC.
 XX Morris DW, Malandro MS;
 XX WPI; 2004-543781/52.
 DR
 XX
 XX New isolated cancer associated nucleic acids comprising at least 10
 PT contiguous nucleotides; useful for diagnosing, preventing and/or treating
 PT cancers such as leukemia and lymphoma.
 XX
 XX Claim 1; SEQ ID NO 792; 199pp; English.
 PS
 XX The present invention relates to cancer associated sequences (AD097025-
 CC AD098004). The sequences are useful for the diagnosis, prevention and/or
 CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence
 CC data for this patent did not form part of the printed specification, but
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 1262 AA;

Query Match 6.6%; Score 96; DB 8; Length 1262;
 Best Local Similarity 22.5%; Pred. No. 3.5;
 Matches 63; Conservative 31; Mismatches 86; Indels 100; Gaps 13;
 QY 24 DPLISGGPLPLESPSRKLTSLSYDP-----TVPSPPMKRARKLGNLTT 70
 DB 912 EPL-----PLPPEEEBELTSPVSOEPKRSABLSDGKATADEKKTK-GRKR----- 959
 QY 71 IKHPEIIGVLPEEDYT-----RADEEPFGQRPFG-RPRKMPHESSTSLMESP 119
 DB 960 -KTKSSSRIDTMQGYTMGTQWPLQQAQOQPGVIVSPYNSRAYPAHSSPALMERL 1018
 QY 120 RKTMPDSKIMPELRKPFEMTAGREEEYSIGRAWVKGHNNNEYEP-----TKAORTDY 174
 DB 1019 RQLQOQPSGYVOQASPYLQPVAG-----SQRLNHQALQOSPLVGGGIDAVLTPA 1068
 QY 175 APNLAVDYLACREIHRMPRPDSKIPELPIVPSRIDEFDATVDPRYETDLKNEYIRHMKQV 234
 DB 1069 HENL-----PSVPLPDQDPRP-----RQOQVHQOQRL 1095
 QY 235 KKGWCAHQRRRTAPARSIALLINKIYQGESKTVBOALGL 274
 DB 1096 LQ-----MOOPQOAPQPO-----QP--SQTQCALGL 1120

RESULT 6
 AAR48994
 ID AAR48994 standard; protein; 531 AA.
 XX
 XX AAR48994;
 AC
 XX 25-MAR-2003 (revised)
 DT 04-SEP-1994 (first entry)
 XX
 DE Human glycosyltransferase Gnt-III protein.
 XX
 KW Glycosyltransferase; enzyme; cancer diagnosis; ss.
 XX Homo sapiens.
 OS
 XX EP585083-A1.
 PN

PD 02-MAR-1994.
 XX
 XX 20-AUG-1993; 93EP-00306628.
 PF
 XX
 PR 21-AUG-1992; 92JP-00243984.
 XX
 XX (TAKI) TAKARA SHUZO CO LTD.
 XX
 XX Nishikawa A, Taniguchi N;
 PI
 XX WPI; 1994-067563/09.
 DR
 XX N-PSDB; AAQ57973.
 DR
 XX
 XX New gene for human glycosyltransferase Gnt-III - and related vectors and
 PT transformed cells; useful in diagnosis of cancer.
 PT
 XX
 XX Claim 3; Page 10; 14pp; English.
 PS
 XX This glycosyltransferase protein is human UDP-N-acetylglucosamine; beta-D
 CC -mannoside-beta-4N-acetylglucosaminyltransferase. It may be expressed
 CC recombinantly in host cells and used in cancer diagnosis. (Updated on 25-
 CC MAR-2003 to correct PN field.)
 CC
 XX
 SQ Sequence 531 AA;

Query Match 6.5%; Score 95.5; DB 2; Length 531;
 Best Local Similarity 22.6%; Pred. No. 1.1;
 Matches 67; Conservative 27; Mismatches 90; Indels 113; Gaps 17;
 QY 6 PLAEFLPEODGRARQNDPLISGGL-PLSPSRKLTSLSYDPTVPSPPMKRARK 63
 DB 53 PVTQASPEFGGDLPLT-PLYSHSPLOPL-PPSKAEBELHRVDLVPDYTEYFVRTK 110
 QY 64 LGNLLTTIKHPEIIGVLPEEDYTADDEEPFGQRPGRPRKMPHESSTSLMESPRKTM 123
 DB 111 AG-----GVCFPKPTKMLERP-----PPGRPEEKPEGANGSARPPRYLL 151
 QY 124 TRDSKIMFELRGKPFEMTAGREEEYSLGRAWK-----GHANNEYPTKAORTDYAPNL 178
 DB 152 SAR-----ERTGR-----GARRKWEVCVCLPGW-----HGpsc 180
 QY 179 AVDYLAEREIHRMPRPDSKIPELPIVPSRI-----DEFD----- 212
 DB 181 GVPFTV--QYSNLPTEKRLVPR--EYPRRVINAINVNHFDLIDVRFHGLGVDVAFVVC 236
 QY 213 -----ATVDP--YETDLKN--EYTRMKQV-----KKGWCAHQRRRT 246
 DB 237 ESNFTAYGEPRPLKREMLTNGTFEYIRH-KYLIVFLDHFPPGGRQDGIADYIART 292

RESULT 7
 AAR97615
 ID AAR97615 standard; protein; 531 AA.
 XX
 XX AAR97615;
 AC
 XX 21-NOV-1996 (first entry)
 DT
 XX Rat N-acetylglucosaminyl transferase-III, inhibitor of metastasis.
 DE
 XX N-acetylglucosaminyl transferase-III; metastasis inhibitor; cancer;
 KW neoplasia; invasive.
 XX
 OS Rattus sp.
 XX
 XX JP08109139-A.
 PN
 XX 30-APR-1996.
 PD
 XX 12-OCT-1994; 94JP-00271802.
 PF
 XX 12-OCT-1994; 94JP-00271802.
 PR
 XX

Human; gene expression; single exon probe; microarray;
alternative splicing event; genomic alteration.
Homo sapiens.
US2003194704-A1.
16-OCT-2003.
03-APR-2002; 2002US-00029386.
03-APR-2002; 2002US-00029386.
(PENN/) PENN S G.
(RANK/) RANK D R.
(HANZ/) HANZEL D K.
Penn SG, Rank DR, Hanzel DK.
WPI; 2004-119264/12.
New human genome-derived single exon nucleic acid probes useful for human
gene expression analysis, for identifying or characterizing alternative
splicing events, for assessing genomic alterations or as tools for
surveying tissues.
Claim 45; SEQ ID NO 34087; 80pp; English.
The invention relates to a nucleic acid probe for measuring human gene
expression, comprising any of the 27,400 fully defined nucleotide
sequences in the specification, or their complements or fragments, and
encoding at least 8 amino acids of any of the 6888 amino acid sequences
fully defined in the specification. The probe is a single exon probe that
hybridizes under high stringency conditions to a nucleic acid molecule
expressed in human cells or tissues. Also included are a spatially-
addressable set of single exon nucleic acid probes for measuring human
gene expression (comprising a plurality of single exon nucleic acid
probes cited above, where each of the plurality of probes is separately
and addressably isolatable or amplifiable from the plurality), a single
exon microarray for measuring human gene expression, a method of
measuring human gene expression, a vector comprising the single exon
probe cited above, an ORF-encoded peptide comprising at least 8
contiguous amino acids of any of the above-mentioned amino acid
sequences (optionally with conservative amino acid substitutions), an
isolated antibody that binds specifically to a peptide cited above, a
method of selling and/or licensing single exon probes or microarrays to
a customer desiring to measure gene expression, a method of providing
human gene expression data by subsequence, a method of providing
storage medium which contains a database having a plurality of records
(each record including data on the expression of a single exon probe
cited above. The probe, methods and apparatus are useful in gene
expression analysis. The probes may be used as tools for surveying
tissues to detect the presence of expressed messages that contain their
specific exon, or in constructing genome-derived single exon microarrays.
In addition, the probes are used in identifying and characterizing
alternative splicing events, in detecting and characterizing gross
alterations in the genomic locus that includes their exon, in assessing
smaller genomic alterations, in priming the synthesis of nucleic acids,
or in expressing the ORF-encoded peptide. The present sequence is a human
single exon probe protein of the invention. Note: The sequence data for
this patent did not form part of the printed specification, but was
obtained in electronic format directly from USPTO at
segdata.uspto.gov/sequence.html?DocID=20030194704

Query Match 6.5%; Score 95.5; DB 8; length 533;
Best Local Similarity 22.6%; Pred. No. 1.1;
Matches 67; Conservative 27; Mismatches 90; Indels 113; Gaps 17;
6 PLAEFLLEDGDRNRQNDPLISGSP--PLESPSRKUTLSLYDPTVPESDMMKPKARK 63

Db	55	PVTFQASPEEGGPDLLNT-PLYSHSPLLQPL-PPSKAAEELHAWDLVLPEDTTEYFVTRK 112
Oy	64	LGNLLTIKHKHPSEILIGVLPEDYTRADEEBGRGPPRGPRKMPRHSSSTLSMESPRKTM 123
Db	113	AG-----GVCFKPGTKMLER-----PPGRPEERPEAGSSARPPRYLL 153
Oy	124	TRDSKIMFELRGKPFEMIAGRFEEBSYSLGAWTK-----GHMNEYEPKIAQRTDYAPNL 178
Db	154	SAR-----ERTGGR-----GARRKWEVCVCLPBM-----HGpsc 182
Oy	179	AVDYLAEREIHRMPRPDKSIPELPIVPSRI-----DEPF----- 212
Db	183	GVPTIV--QVSNLPTKERLVPR--EYPRRYVIAINVNHEFDLDVRFHFLGDVYDAFVVC 238
Oy	213	-----ATVDR--YETDLKN--EYIRMKOV-----KKGWCAHQRRRT 246
Db	239	ESNFTAYGEPRLPKFREMLTNGTFEYIRH-KVLYVFLDHPFGGRQDDGIADLYLRT 294
RESULT 10		
ID	ADD68089	
XX	ADD68089	standard, protein, 544 AA.
XX	AC	ADD68089;
XX	DT	15-JAN-2004 (first entry)
XX	DE	Human galactosyltransferase GntTIII SEQ ID NO:6.
XX	KW	human, galactosyltransferase; transmembrane; glycosyltransferase;
XX	KM	glycan processing.
XX	OS	Homo sapiens.
XX	PN	MO2003078637-A2.
XX	PD	25-SEP-2003.
XX	PF	18-MAR-2003; 2003W0-IB001626.
XX	PR	19-MAR-2002; 2002US-0365735P.
XX	PA	(PLAN-) PLANT RES INT BV.
XX	PI	Bakker HAC, Florack DEA, Bosch HJ, Rouwendal GJA;
XX	XX	
XX	DR	WPI: 2003-902819/82.
XX	DR	N-PSDB; ADD68088.
XX	PT	New nucleic acid encoding a hybrid enzyme comprising a transmembrane
XX	PT	region of a plant glycosyltransferase and a catalytic region of a
XX	PT	mammalian glycosyltransferase, useful for optimizing glycan processing in
XX	PT	organism.
XX	PS	Disclosure; SEQ ID NO 6; 139pp; English.
XX	CC	The invention relates to a novel nucleic acid encoding a hybrid enzyme
XX	CC	comprising a transmembrane region of a plant glycosyltransferase and a
XX	CC	catalytic region of a mammalian glycosyltransferase. The nucleic acid
XX	CC	molecules are useful for optimising glycan processing in an organism so
XX	CC	that a glycoprotein having complex bi-antennary glycans and containing
XX	CC	galactose residues on both arms and which are devoid of xylose and fucose
XX	CC	can be obtained. The present sequence is used in the exemplification of
XX	CC	the invention.
XX	XX	
XX	XX	Sequence 544 AA;

Query Match 6.5%; Score 95.5; DB 7; Length 544;
Best Local Similarity 22.6%; Pred. No. 1.2;
Matches 67; Conservative 27; Mismatches 90; Gaps 17;
QY 6 PLAEFLLPEDGDRNARNDDPLTGGPI -PLESPSRKLTSLSYDPTVPSPDMKFAKR 63

```
Db 55 PVTPOASEPGGDLIRT-PLYSHSPILQPL-PPSKAAEELHRVDLVLPEDTTEYFVRTK 112
Qy 64 LGNLTITTKHHPSELIGLVPEDYTRADEPBGQGRPPRPRKMRPHESSTISMEPRKTM 123
Db 113 AG-----GVCFKPGTKMLERP-----PPGRPEKPEGANGSSARRPPRYLL 153
Qy 124 TRDSKIMFELNGKPPREMIAGRFEEYSLSGRAMVK-----GHMNNYEPIKQRTDYANL 178
Db 154 SAR-----ERTGR-----GARRKMWECVCLPGM-----HGFSK 182
Qy 179 AVDYLAEREIHRMPRPDKSIPELPIVPSRI-----DEPD----- 212
Db 183 GVPTIV--QYGNLPTFKERLVPR--EVPRRVAINAINVNHFFLLDVRFHFLGDVDAFYVC 238
Qy 213 ----ATVDP--YETDLKN---EYIRHWKQV-----KKQWCHQRRT 246
Db 239 ESNFTAYGEPRPLKREMLTNGTFEYIRH-KVLVYFLDHPFPGGRQDGMIDDYILRT 294
```

RESULT 11

ABR83694
ID ABR83694 standard; protein; 544 AA.

XX ABR83694;

DT 12-FEB-2004 (first entry)

XX Human GntIII enzyme.

XX GntIII; N-acetylglucosaminyltransferase; transgenic; glycoprotein; human;

XX enzyme.

XX Homo sapiens.

XX WO2003078614-A2.

XX 25-SEP-2003.

XX 18-MAR-2003; 2003WO-1B001562.

XX 19-MAR-2002; 2002US-0365769P.

XX 26-MAR-2002; 2002US-0368047P.

XX (PLAN-) PLANT RES INT BV.

XX Bakker HAC, Florack DEA, Bosch HJ;

XX WPI; 2003-779132/73.

XX N-PSDB; ACF58335.

XX New plant host cell system for producing a desired glycoprotein comprises

XX PT a mammalian N-acetylglucosaminyltransferase (GntIII) enzyme, a nucleic

XX PT acid sequence encoding the enzyme, or a vector comprising the nucleic

XX PT acid sequence.

XX Example 1; Fig 4A-B; 122pp; English.

XX The invention relates to a plant host cell system comprising a mammalian

XX CC UDP-N-acetylglucosamine; beta-D-mannoside beta(1,4)-N-

XX CC acetylglucosaminyltransferase (GntIII) enzyme, a nucleic acid sequence

XX CC encoding a mammalian GntIII protein, or a vector comprising the GntIII

XX CC nucleic acid. The plant host system is useful in producing a desired

XX CC glycoprotein or its functional fragment. The glycoprotein or its

XX CC functional fragment may be used for the production of a pharmaceutical

XX CC composition. The present sequence represents a human GntIII enzyme

XX SQ Sequence 544 AA;

Query Match 6.5%; Score 95.5; DB 7; Length 544;

Best Local Similarity 22.6%; Pred. No. 1.2;

Matches 67; Conservative 27; Mismatches 90; Indels 113; Gaps 17;

Qy 6 PLAEFLPEDDGRNARQNDPLISGGL--PLESPSKLTSLSYDPTVPESPDMKFAKR 63

```
Db 55 PVTPOASEPGGDLIRT-PLYSHSPILQPL-PPSKAAEELHRVDLVLPEDTTEYFVRTK 112
Qy 64 LGNLTITTKHHPSELIGLVPEDYTRADEPBGQGRPPRPRKMRPHESSTISMEPRKTM 123
Db 113 AG-----GVCFKPGTKMLERP-----PPGRPEKPEGANGSSARRPPRYLL 153
Qy 124 TRDSKIMFELNGKPPREMIAGRFEEYSLSGRAMVK-----GHMNNYEPIKQRTDYANL 178
Db 154 SAR-----ERTGR-----GARRKMWECVCLPGM-----HGFSK 182
Qy 179 AVDYLAEREIHRMPRPDKSIPELPIVPSRI-----DEPD----- 212
Db 183 GVPTIV--QYGNLPTFKERLVPR--EVPRRVAINAINVNHFFLLDVRFHFLGDVDAFYVC 238
Qy 213 ----ATVDP--YETDLKN---EYIRHWKQV-----KKQWCHQRRT 246
Db 239 ESNFTAYGEPRPLKREMLTNGTFEYIRH-KVLVYFLDHPFPGGRQDGMIDDYILRT 294
```

RESULT 12

ABO79565
ID ABO79565 standard; protein; 530 AA.

XX ABO79565;

DT 29-JUL-2004 (first entry)

XX Pseudomonas aeruginosa polypeptide #11740.

XX Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.

XX Pseudomonas aeruginosa.

XX US6551795-B1.

XX 22-APR-2003.

XX 18-FEB-1999; 99US-00252991.

XX 18-FEB-1998; 98US-0074788P.

XX 27-JUL-1998; 98US-0094190P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;

XX WPI; 2003-615309/58.

XX N-PSDB; ABD13136.

XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,

XX PT useful as molecular targets for diagnostics, prophylaxis and treatment of

XX PT pathological conditions resulting from bacterial infection.

XX Disclosure; SEQ ID NO 28311; 455pp; English.

XX The invention relates to Pseudomonas aeruginosa polypeptides and the

XX CC polynucleotides encoding them. The sequences are useful in diagnostics and

XX CC therapy of pathological conditions, as molecular targets for diagnostics,

XX CC prophylaxis and treatment of pathological conditions resulting from a

XX CC bacterial infection, for evaluating a compound, such as a polypeptide,

XX CC for the ability to bind a P. aeruginosa nucleic acid, as components of

XX CC effective antibacterial targets, as targets for antibacterial drugs,

XX CC including anti-P. aeruginosa drugs, as templates for recombinant

XX CC production of P. aeruginosa-derived peptides or polypeptides, as target

XX CC components for diagnosis and/or treatment of P. aeruginosa-caused

XX CC infection, and in detection of P. aeruginosa sequences or other sequences

XX CC of Pseudomonas species using bioclip technology. Sequences ABO67826-

XX CC ABO64386 represent P. aeruginosa polypeptides of the invention. Note: The

XX CC sequence data for this patent did not form part of the printed

XX CC specification but was obtained in electronic format from USPTO at

XX CC seqdata.uspto.gov/sequence.html

[illegible]

RESULT 13
 ABB70419
 ID ABB70419 standard; protein; 1161 AA.
 AC ABB70419;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 38049.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KM pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 WO200171042-A2.
 PN
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US009231.
 XX
 PR 23-MAR-2000; 2000US-0191637P.
 XX
 PR 11-JUL-2000; 2000US-00614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 DR WPI; 2001-656860/75.
 XX
 DR N-PSDB; ABL14522.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 XX
 DT Disclosure; SEQ ID NO 38049; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention
 CC is useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly

[illegible]

XX	ABU05565
XX	ABU05565 standard; protein; 410 AA.
AC	ABU05565;
XX	
DT	08-APR-2003 (first entry)
XX	
DE	M. tuberculosis and M. leprae marker protein #216.
XX	
KM	Mycobacterioses; survival; virulence; protective antigen; vaccine;
KW	mycobacterial disease; tuberculosis; leprosy.
XX	
OS	Mycobacterium tuberculosis.
XX	Mycobacterium leprae.
FN	NO200274903-A2.
PD	26-SEP-2002.
XX	
PF	22-FEB-2002; 2002WC-IB001973.
XX	
PR	22-FEB-2001; 2001US-0270123P.
PA	(INSP) INST PASTEUR.
PI	Cole S;
DR	WPI; 2002-759885/82.
PT	Identifying and selecting genes for survival or virulence of mycobacteria
PT	by a comparative genomic analysis of the sequences of Mycobacterium
PS	tuberculosis and M. leprae.
XX	
XX	Claim 17; Page 394-396; 874pp; English.
XX	
CC	This invention relates to a novel method for identifying essential genes
CC	for survival or virulence of mycobacteria species. The method comprises
CC	aligning the genomic sequence of a first mycobacterium species on a
CC	genomic sequence of a second mycobacterium species and selecting a
CC	polynucleotide sequence that is highly conserved in both genomes with no
CC	correspondence in other bacterial genomic sequences and that corresponds to
CC	an essential gene for the survival or virulence of mycobacterium species.
CC	The method of the invention is useful for detecting M. tuberculosis or M.

THIS PAGE BLANK (USPTO)